Approximation of Integrals over asymptotic sets

with applications to Statistics and Probability

Philippe Barbe

Philippe Barbe

APPROXIMATION
of
INTEGRALS
over
ASYMPTOTIC SETS

with applications to

Statistics and Probability

Philippe Barbe
CNRS
90 Rue de Vaugirard
75006 PARIS
FRANCE

Contents

1. Introduction

Computing integrals is a basic need in many areas of mathematics and applied sciences. It is a common experience that some integrals can be calculated "by hand", obtaining a "closed formula", while others cannot be reduced to a simple expression. This fact is at the origin of numerous approximation techniques for integrals, such as quadratures, series expansions, Monte-Carlo methods, etc. The purpose of these notes is to start what seems to be a new line of investigations in studying the behavior of integrals of the form

$$\int_A f(x)\mathrm{d}x\,, \qquad A\subset\mathbb{R}^d\,, \tag{1.1}$$

for sets A far away from the origin. In this integral, $\mathrm{d}x$ is the Lebesgue measure on $\mathbb{R}^d$, and the function f is integrable over $\mathbb{R}^d$.

Such integral is in general very difficult to compute numerically. Indeed, in interesting applications d is between 10 and 50 say, and quadrature methods are essentially out of considerations. On the other hand, as A is far away from the origin, the integral is small — the interesting range for some applications is when the integral is of order 10^{-2}, 10^{-3}, or even smaller — and standard Monte-Carlo methods fail to provide a good evaluation at a cheap cost.

The motivation for this study comes mainly from applications in statistics and probability theory. However, the techniques developed should be useful in other areas of mathematics as well, wherever such integrals arise.

Before going further, let us show some consequences of our main approximation result in probability and statistics. We hope that the reader will find in these examples good motivation to continue reading.

Consider the following two density functions on $\mathbb{R}$,

$$\begin{aligned} w_\alpha(x) &= K_{w,\alpha}\exp\bigl(-|x|^\alpha/\alpha\bigr)\,,\\ s_\alpha(x) &= K_{s,\alpha}\Bigl(1+\frac{x^2}{\alpha}\Bigr)^{-(\alpha+1)/2}\,, \end{aligned}$$

where

$$K_{w,\alpha} = \frac{1}{2\alpha^{\frac{1}{\alpha}-1}\Gamma\Bigl(\frac{1}{\alpha}\Bigr)} \quad \text{and} \quad K_{s,\alpha} = \frac{1}{\sqrt{\pi\alpha}} \frac{\Gamma\Bigl(\frac{\alpha+1}{2}\Bigr)}{\Gamma\Bigl(\frac{\alpha}{2}\Bigr)} ;$$

so w_α and s_α integrate to 1 on the real line. The density w_2 is the standard normal distribution. The density s_α is that of a Student distribution. These two functions are very different. The symmetric Weibull-like, w_α, decays exponentially at infinity, while the Student one decays polynomially.

Let us now consider a real $d \times d$ matrix $C = (C_{i,j})_{1\leqslant i,j\leqslant d}$ with at least one positive term on its diagonal. We write $\langle \cdot\,, \cdot\rangle$ the standard inner product on $\mathbb{R}^d$, and consider the domain

$$A_t = \{\, x \in \mathbb{R}^d : \langle Cx, x\rangle \geqslant t \,\} .$$

As t tends to infinity, it is easy to see that the set $A_t = \sqrt{t}A_1$ pulls away from the origin since A_1 does not contain 0. We will explain the following in section 8.1. If $\alpha \neq 2$, then there exists a function — rather explicit, and in any case computable — $c(\cdot)$ of α, d and C, such that

$$\int_{A_t} \prod_{1\leqslant i\leqslant d} w_\alpha(x_i)\mathrm{d}x_1 \ldots \mathrm{d}x_d \sim c(\alpha, d, C)e^{-t^{\alpha/2}c_2} t^{-(\alpha-2)\frac{d}{4}-\frac{\alpha}{2}} ,$$

as t tends to infinity. When $\alpha = 2$, the asymptotic behavior of the integral is different. Namely, if k is the dimension of the eigensubspace associated to the largest eigenvalue λ of $C^{\mathrm{T}} + C$,

$$\int_{A_t} \prod_{1\leqslant i\leqslant d} w_2(x_i)\mathrm{d}x_1 \ldots \mathrm{d}x_d \sim c(2, d, C)e^{-t/\lambda} t^{(k-1)/2} ,$$

as t tends to infinity. Comparing the formula for α equal or different from 2, we see that the power of t, namely $(\alpha-2)d/4 - \alpha/2$ for α different from 2 and $(k-1)/2$ for α equal 2, has a discontinuity in $\alpha = 2$, whenever k is different than 2. Moreover, we will see in section 8.2 that

$$\int_{A_t} \prod_{1\leqslant i\leqslant d} s_\alpha(x_i)\mathrm{d}x_1 \ldots \mathrm{d}x_d \sim K_{s,\alpha}\alpha^{(\alpha+1)/2} 2t^{-\alpha/2} \sum_{i:C_{i,i}>0} C_{i,i}^{\alpha/2}$$

as t tends to infinity. These estimates give approximations for the tail probability of the quadratic form $\langle CX, X\rangle$ evaluated at a random vector X having independent components distributed according to w_α or s_α. We will also see what happens when all the diagonal elements of C are negative.

Consider now $d = n^2$ for some positive integer n, and write $\mathrm{M}(n,\mathbb{R})$ for the set of all real $n\times n$ matrices. Let

$$A_t = \{\, x \in \mathrm{M}(n,\mathbb{R}) : \det x \geqslant t \,\}$$

be the set of all $n\times n$ real matrices with determinant at least t. The set A_1 is closed and does not contains the origin. One sees that $A_t = t^{1/n}A_1$ moves away from the origin when t increases. We will prove that, as t tends to infinity,

$$\int_{A_t} \prod_{1\leqslant i,j\leqslant n} w_\alpha(x_{i,j})\mathrm{d}x_{i,j} \sim \begin{cases} c\, e^{-t^\alpha/n} t^{(\alpha(n^2-1)-2n^2)/2n} & \text{if } \alpha\neq 2 \\ c\, e^{-nt^2/2} t^{(n^2-n-2)/2} & \text{if } \alpha = 2 \end{cases}$$

while

$$\int_{A_t} \prod_{1\leqslant i,j\leqslant n} s_\alpha(x_{i,j})\mathrm{d}x_{i,j} \sim c\,\frac{(\log t)^{n-1}}{t^\alpha}\,.$$

This gives a tail estimate of the distribution of the determinant of a random matrix with independent and identically distributed entries from a distribution w_α or s_α. The constant c depends on s_α or w_α as well as on the dimension n; but it does not depend on t and will be explicit.

Possible applications are numerous. We will deal with further interesting examples, such as norms of random matrices, suprema of linear processes, and other related quantities.

Though some specific examples could be derived with other methods — in particular those dealing with the Gaussian density integrated over rather simple sets — we hope that the reader will enjoy having a unified framework to handle all those asymptotic problems. It has the great advantage of providing a systematic approach. It breaks seemingly complicated problems into much more manageable ones. It also brings a much better understanding of how the leading terms come out of the integrals. Through numerical computations we will also see that our approximations are accurate enough to be of practical use in statistics, when dealing with some small sample problems — the type of problem for which there are no systematic tools as far as the author knows, and which originally motivated this work.

Another purpose of these notes is to investigate asymptotics for conditional distributions, and Gibbs conditioning. For instance, consider n points in $\mathbb{R}^n$, whose coordinates are independent and identically distributed random variables. They form a parallelogram. Given, say, the volume of this parallelogram, these points are no longer independent. What is their distribution? In general, this seems to be a very difficult question, except if one stays at a very theoretical level and does not say much. We will show that it is possible to obtain some good approximation of this conditional distribution for large volumes of the parallelogram — mathematically, as the volume tends to infinity.

It is important to realize that conditional distributions are rather delicate objects. To convince the reader of the truth of this assertion, let us consider four elementary examples where everything can be calculated explicitly and easily.

Example 1. Let X, Y be two independent standard normal random variables. We are looking for the distribution of (X,Y) given $X+Y \geqslant t$, for large values of t. A simple probabilistic argument to obtain a limiting distribution is as follows. Let $U = (X+Y)/\sqrt{2}$ and $V = (X-Y)/\sqrt{2}$. The distribution of X given $X+Y \geqslant t$ is that of $(U+V)/\sqrt{2}$ given $U \geqslant t/\sqrt{2}$. One can check by hand that the conditional distribution of U/t given $U \geqslant t/\sqrt{2}$ converges weakly* to a point mass at $1/\sqrt{2}$. Since U and V are independent standard normal, and since (X,Y) is exchangeable conditionally on $X+Y$, the pair $(X,Y)/t$ converges in probability to $(1/2, 1/2)$ conditioned on $X+Y \geqslant t$.

Consequently, if we are given $X+Y \geqslant t$ for large t, we should expect to observe $X/t \approx Y/t \approx 1/2$. Hence, X and Y are both large, and about the same order.

Example 2. Consider the same problem as in example 1, but with X and Y independent and both exponential. For any $\alpha \in (0,1)$,

$$P\{\, X \leqslant \alpha t \,|\, X+Y \geqslant t \,\} = \frac{P\{\, X \leqslant \alpha t \,;\, X+Y \geqslant t \,\}}{P\{\, X+Y \geqslant t \,\}} \, .$$

We can explicitly calculate

$$\begin{aligned} P\{\, X \leqslant \alpha t \,;\, X+Y \geqslant t \,\} &= \int_{x=0}^{\alpha t} \int_{y=t-x}^{\infty} e^{-x} e^{-y} \mathrm{d}x \, \mathrm{d}y = \int_0^{\alpha t} e^{-t} \mathrm{d}x \\ &= \alpha t e^{-t} \, . \end{aligned}$$

On the other hand, $X+Y$ has a gamma distribution with mean 2. Integration by parts yields

$$P\{X+Y \geqslant t\} = \int_t^\infty xe^{-x}\mathrm{d}x = (t+1)e^{-t}\,.$$

It follows that

$$P\{X \leqslant \alpha t \mid X+Y \geqslant t\} = \alpha\frac{t}{t+1}\,.$$

Thus, given $X+Y \geqslant t$, the distribution of X/t converges to a uniform distribution over $[0,1]$. Again, by conditional exchangeability, the same result holds for Y/t. In conclusion, the distribution of $(X,Y)/t$ conditioned on $X+Y \geqslant t$ converges weakly* to a uniform distribution over the segment $\alpha(1,0)+(1-\alpha)(0,1)$ in $\mathbb{R}^2$. So, we do not have the kind of degeneracy that occurs in Example 1. Conditioned on $X+Y \geqslant t$, the random variables X and Y should still be of the same order of magnitude, t, but not necessarily close to each others after the rescaling by $1/t$.

Example 3. Consider the same problem, with now X and Y independent, both having a Cauchy distribution. Elementary calculation shows that

$$P\{X \geqslant s\} = \int_s^\infty \frac{\mathrm{d}x}{\pi(1+x^2)} \sim \int_s^\infty \frac{\mathrm{d}x}{\pi x^2} \sim \frac{1}{\pi s} \qquad \text{as } s \to \infty\,.$$

Moreover, the stability of the Cauchy distribution — see, e.g., Feller, 1971, §II.4 — asserts that $(X+Y)/2$ has also a Cauchy distribution. Consequently,

$$\begin{aligned}\lim_{t\to\infty} P\{X \geqslant \epsilon t\,, Y \geqslant \epsilon t \mid X+Y \geqslant t\}&\\ \leqslant \lim_{t\to\infty} \frac{P\{X \geqslant \epsilon t\,; Y \geqslant \epsilon t\}}{P\{X+Y \geqslant t\}} &= 0\,.\end{aligned}$$

This proves that we cannot have both X and Y too large as t tends to infinity. Moreover, for s positive,

$$P\{X \geqslant t(1+s) \mid X+Y \geqslant t\} \sim \frac{P\{X \geqslant t(1+s)\}}{P\{X+Y \geqslant t\}} \sim \frac{1}{2(1+s)}$$

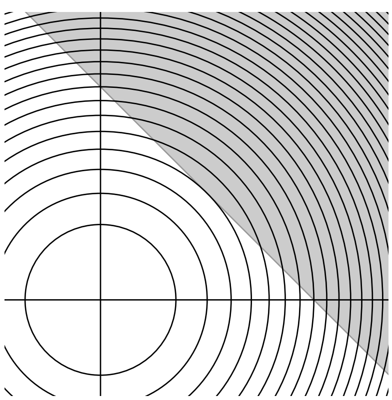
as t tends to infinity. Consequently, given $X+Y \geqslant t$, the random vector $(X,Y)/t$ has a distribution which converges weakly* to $(P_x +$

$P_y)/2$ where P_x (resp. P_y) is a Pareto distribution on the x-axis (resp. y-axis).

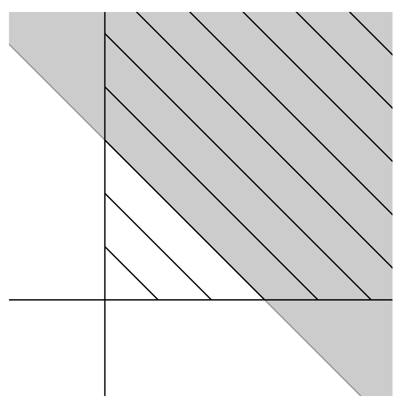
These three examples show that we can obtain very different behaviors, with various degrees of degeneracy for the limiting conditional distributions. But they also show examples of nonconceptual proofs. They provide no insight on what is happening. All the arguments are rather ad hoc. However, it is intuitively clear that the limiting conditional distribution has to do with the tail of the distribution of (X, Y). Specifically, if we write $f(x, y)$ for the density of (X, Y) and $A_t = \{ (x, y) : x + y \geqslant t \}$, we are led to consider expressions of the form

$$\frac{\displaystyle\int_{A_t \cap \{ x : x \leqslant ct \}} f(x, y) \,\mathrm{d}x \,\mathrm{d}y}{\displaystyle\int_{A_t} f(x, y) \,\mathrm{d}x \,\mathrm{d}y} .$$

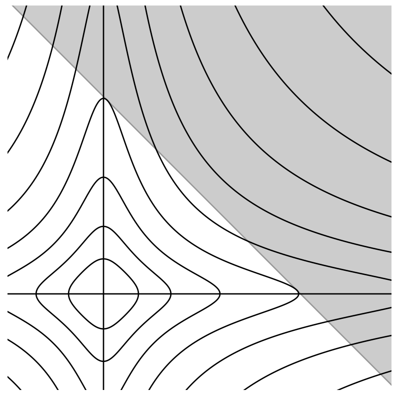
When looking at $\int_{A_t} f(x, y) \,\mathrm{d}x \,\mathrm{d}y$, examples 1, 2, 3 have very different features. Intuitively, the behavior of the integral has to do with how the set A_t lies in $\mathbb{R}^2$ compared to the level sets of $f(x, y)$. Examples 1, 2 and 3 correspond to the following pictures.



Example 1



Example 2



Example 3

In the three pictures, we see the gray shaded set A_t, its boundary, and the level sets

$$\Lambda_c = \{ (x,y) : f(x,y) = e^{-c} \}$$

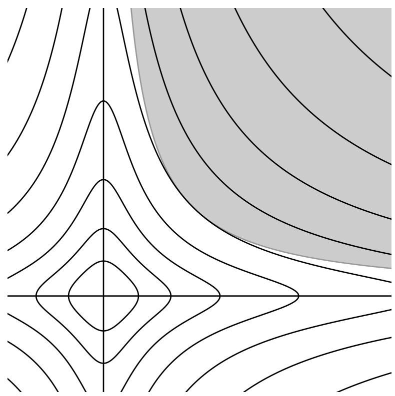
for $c = 1, 2, 3, \ldots$ The fact that the sets Λ_c are getting closer and closer as c tends to infinity in example 1 expresses the very fast decay of the Gaussian distribution. At the opposite extreme, in example 3, the level sets are further and further apart because the distribution has a subexponential decay. These pictures show why in examples 1 and 3 we obtained some form of degeneracy in the limiting distribution. In example 1, the density function is maximal on the boundary of A_t at one point whose two coordinates are equal. In example 3, the density is maximal at two points on the boundary of A_t. These two points have one coordinate almost null, while the other coordinate is of order t. But one should be careful, as rough pictures may not give the right result if they are read naively in more complicated situations. Our fourth example is still elementary and illustrates this point.

Example 4. Consider (X, Y) where X and Y are independent, both having a Cauchy distribution with density s_1. We are now interested in the tail distribution of XY and in the distribution of (X, Y) given $XY \geqslant t$ for large t. By symmetry, it is enough to consider the same distribution but adding the conditioning $X \geqslant 0$ and $Y \geqslant 0$.

On $X, Y > 0$, set $U = \log X$ and $V = \log Y$. The density of U is

$$f_U(u) = e^u s_1(e^u) = \frac{e^u}{\pi(1+e^{2u})} \sim \frac{e^{-u}}{\pi} \qquad \text{as } u \to \infty\,.$$

Thus, the level sets of the density of (U,V) far away from the origin look like those of the exponential distribution of example 2. Hence, when studying the distribution of U given $U+V \geqslant \log t$ — i.e., X given $XY \geqslant t$ — we could expect to have a behavior similar to example 2. However, the picture looks as follows.



Example 4

In particular, write $A_t = \{\,(x,y) : xy \geqslant t\,\}$ for the set on which we integrate the density of (X,Y). The boundary ∂A_t has a unique contact point with the maximal level set of the density intersecting A_t, where by maximal level set we mean Λ_c such that $\Lambda_{c+\epsilon} \cap A_t$ is nonempty if ϵ is positive, and is empty if ϵ is negative. In that aspect, we are close to examples 1 or 3. This example 4 will show us two more things. First we need to take into consideration the decay of the density — which explains why we will not have degeneracy as in examples 1 or 3. Second, it shows that whichever result we can prove will have a lot to do with the shape of the domain considered and how this domain pulls away from the level sets of the density.

To see what happens in this fourth example is still rather easy. We first calculate

$$\begin{aligned} P\{\,XY \geqslant t\,\} &= \frac{2}{\pi^2} \int_{xy\geqslant t\,,\,x\geqslant 0\,,\,y\geqslant 0} \frac{\mathrm{d}y}{1+y^2}\frac{\mathrm{d}x}{1+x^2} \\ &= \frac{2}{\pi^2}\int_0^\infty \Bigl(\frac{\pi}{2} - \arctan\frac{t}{x}\Bigr)\frac{\mathrm{d}x}{1+x^2}\,. \end{aligned} \tag{1.2}$$

It is true that

$$\frac{\pi}{2} - \arctan u = \int_u^\infty \frac{\mathrm{d}x}{1+x^2} \sim \frac{1}{u} \qquad \text{as } u \to \infty\,.$$

However, replacing $(\pi/2) - \arctan(t/x)$ by x/t in (1.2) leads to a divergent integral. So, we need to be careful. Let η be a positive real number. There exists a positive ϵ such that for any $u > 1/\epsilon$,

$$\frac{\pi}{2} - \arctan u \in \Big[\frac{1-\eta}{u}, \frac{1+\eta}{u}\Big].$$

Consequently,

$$\int_0^{\epsilon t} \Big(\frac{\pi}{2} - \arctan\frac{t}{x}\Big)\frac{\mathrm{d}x}{1+x^2}$$
$$\begin{cases} \leqslant (1+\eta)\displaystyle\int_0^{\epsilon t} \frac{x\mathrm{d}x}{t(1+x^2)} = \frac{1+\eta}{2t}\log(1+\epsilon^2t^2) \\ \geqslant (1-\eta)\displaystyle\int_0^{\epsilon t} \frac{x\mathrm{d}x}{t(1+x^2)} = \frac{1-\eta}{2t}\log(1+\epsilon^2t^2)\,. \end{cases}$$

Moreover,

$$0 \leqslant \int_{\epsilon t}^{\infty}\Big(\frac{\pi}{2} - \arctan\frac{t}{x}\Big)\frac{\mathrm{d}x}{1+x^2} \leqslant \frac{\pi}{2}\int_{\epsilon t}^{\infty}\frac{\mathrm{d}x}{x^2} = \frac{\pi}{2\epsilon t}\,.$$

Therefore, using $\log(1+\epsilon^2t^2) \sim 2\log t$ as t tends to infinity, we have

$$P\{\,XY \geqslant t\,\} \sim \frac{2}{\pi^2}\frac{\log t}{t} \qquad \text{as } t \to \infty\,.$$

Next, for $\alpha \in (0,1)$, the same argument shows that

$$\begin{aligned} P\{\,XY \geqslant t\,; X \geqslant t^\alpha\,\} &= \frac{1}{\pi^2}\int_{t^\alpha}^{\infty}\int_{t/x}^{\infty}\frac{\mathrm{d}y}{1+y^2}\frac{\mathrm{d}x}{1+x^2} \\ &= \frac{1}{\pi^2}\int_{t^\alpha}^{\infty}\Big(\frac{\pi}{2} - \arctan\frac{t}{x}\Big)\frac{\mathrm{d}x}{1+x^2} \\ &\sim \frac{1}{\pi^2}\int_{t^\alpha}^{\epsilon t}\frac{x}{t}\frac{\mathrm{d}x}{1+x^2} \\ &= \frac{1}{t\pi^2}\big(\log\epsilon t - \log t^\alpha + o(1)\big) \\ &\sim \frac{\log t}{t\pi^2}(1-\alpha) \qquad \text{as } t \to \infty. \end{aligned}$$

Consequently, we obtain

$$P\{\,X \geqslant t^\alpha \,|\, XY \geqslant t\,\} = \frac{P\{\,X \geqslant t^\alpha\,; XY \geqslant t\,\}}{P\{\,XY \geqslant t\,\}} \sim \frac{1-\alpha}{2}\,.$$

It follows that the distribution of $(X,Y)/t$ given $XY \geqslant t$ converges weakly* to $\delta_{(0,0)}$, showing a degeneracy like in example 1, but of a different nature. A linear normalization is not suitable, so we must use a logarithmic scale. Writing $X = \epsilon_1 e^U$ and $Y = \epsilon_2 e^V$ with ϵ_1 and ϵ_2 taking values $+1$ or -1 with probability $1/2$, we proved that the distribution of $(U,V)/\log t$ given $XY \geqslant t$ converges weakly* to a uniform distribution over $\{\,(s, 1-s) : s \in [\,0,1\,]\,\}$, which is very similar to example 2.

Another interesting feature about examples 1–4 is that they use some rather specific arguments in each case, since everything could be calculated quite explicitly. When looking at more complicated distributions, or at more complicated conditioning, or in higher dimensions, we cannot rely so much on elementary intuition. One of the main goals of the present work is to give a systematic procedure for doing the calculation. The key point is that we will be able to transform the problem into one in asymptotic differential geometry, i.e., a conjunction of asymptotic analysis and differential geometry. In practice, the approximation of the conditional distribution will boil down to a good understanding of the contact between ∂A_t and the level sets of the density. Having to deal with purely geometrical quantities will be helpful because differential geometric arguments will not depend on the dimension, and also since differential geometric quantities, such as curvatures, can be computed from different formulas, using very different parameterizations. Whichever parameterization is the most convenient will be used. In some sense, this is very much like changing variables in integrals, and we hope to convince the reader that once it has been learned, it considerably simplifies the analysis. The disadvantage is that it requires investing some time in learning the basics of differential geometry and asymptotic analysis — but can we really hope to solve every problem with elementary calculus?

At this stage, I urge the reader to look at the conditional distribution of the parallelogram given its volume mentioned in the introduction. In terms of linear algebra, the question amounts to looking at the conditional distribution of a random matrix with independent and identically distributed coefficients, given that its determinant is large. For 2×2 matrices, we are already in $\mathbb{R}^4$, since we have 4 coefficients. The condition that the determinant is larger than t determines a subset whose boundary is of dimension 3, and it is impossible to visualize it. Such a simple example already shows the need for systematic

methods relying as little as possible on intuition.

Let us now say a few words on how our results are connected to the classical Laplace method. They can be viewed as a Laplace method with an infinite dimensional parameter.

In order to discuss this point, and because some understanding of Laplace's method may be helpful in reading the next pages, let us state and prove the following elementary result. Writing $I = -\log f$, consider an integral of the form $\int_A e^{-\lambda I(x)} \mathrm{d}x$.

1.1. THEOREM. *(Laplace's approximation)* *Let A be a compact set in $\mathbb{R}^d$ and I be a twice differentiable function on $\mathbb{R}^d$. Assume that I has a minimum on A at a unique interior point a_*, and that $\mathrm{D}^2 I(a_*)$ is definite. Then,*

$$\int_A e^{-\lambda I(x)} \mathrm{d}x \sim \frac{(2\pi)^{d/2}}{\big(\det \mathrm{D}^2 I(a_*)\big)^{1/2}} \lambda^{d/2} e^{-\lambda I(a_*)} \qquad \textit{as } \lambda \to \infty\,.$$

Proof. Let ϵ be a number between 0 and 1/2. Since a_* is in the interior of A and I is twice differentiable, we can find an open neighborhood U of a_* on which

$$\begin{aligned}\frac{1-\epsilon}{2} \langle \mathrm{D}^2 I(a_*)(x-a_*)\,,\, x-a_*\rangle &\leqslant I(x) - I(a_*)\\ &\leqslant \frac{1+\epsilon}{2}\langle \mathrm{D}^2 I(a)(x-a_*), x-a_*\rangle\,.\end{aligned}$$

Moreover, there exists a positive η such that $I(x) \geqslant I(a_*) + \eta$ on $A \cap U^{\mathrm{c}}$. Consequently,

$$\begin{aligned}\int_A e^{-\lambda I(x)} \mathrm{d}x \leqslant e^{-\lambda I(a_*)} \int_{A\cap U} \exp\Big(\frac{1-\epsilon}{2}\langle \mathrm{D}^2 I(a_*)(x-a_*), x-a_*\rangle\Big)\mathrm{d}x\\ + e^{-\lambda(I(a_*)+\eta)}|A\cap U^{\mathrm{c}}|\,.\end{aligned}$$

The change of variable $h = \sqrt{\lambda(1-\epsilon)}(x-a_*)$ yields to the upper bound

$$\begin{aligned}e^{-\lambda I(a_*)} \int_{\sqrt{\lambda(1-\epsilon)}(A\cap U - a_*)} \exp\Big(-\frac{1}{2}\langle \mathrm{D}^2 I(a_*)h, h\rangle\Big)\mathrm{d}h\, \lambda^{-d/2}(1-\epsilon)^{-d/2}\\ + e^{-\lambda(I(a_*)+\eta)}|A\cap U^{\mathrm{c}}|\,.\end{aligned}$$

Since a_* is an interior point of A, the set $\sqrt{\lambda(1-\epsilon)}(A \cap U - a_*)$ expands to fill $\mathbb{R}^d$ as λ tends to infinity. Therefore,

$$\int_A e^{-\lambda I(x)} \mathrm{d}x \leqslant \frac{\lambda^{-d/2}}{\big(\det \mathrm{D}^2 I(a_*)\big)^{1/2}} e^{-\lambda I(a_*)} (1-2\epsilon)^{-d/2}$$

for λ large enough. A similar argument leads to the lower bound where the term $(1-2\epsilon)^{-d/2}$ is replaced by $(1+2\epsilon)^{-d/2}$. Since ϵ is arbitrary, the result follows. ■

How is our integral $\int_A e^{-I(x)} \mathrm{d}x$ related to Laplace's method? If $A = tA_1$ for some fixed set A_1, and if I is a homogeneous function of degree α, we see that

$$\int_{tA_1} e^{-I(x)} \mathrm{d}x = t^d \int_{A_1} e^{-I(ty)} \mathrm{d}y = t^d \int_{A_1} e^{-t^\alpha I(y)} \mathrm{d}y \,. \tag{1.3}$$

On one hand Laplace's method is directly applicable provided I has a unique minimum on A_1. On the other hand, the integral on the left hand side of (1.3) is like (1.1). Having a good estimate for (1.1) will yield an estimate of (1.3).

When studying (1.1), one can try to argue as in Laplace's method. Writing $I(A)$ for the infimum of I over A, we obtain

$$\int_A e^{-I(x)} \mathrm{d}x = e^{-I(A)} \int_A e^{-(I(x)-I(a_*))} \mathrm{d}x \,.$$

We can hope to have a quadratic approximation for $I(x) - I(a_*)$. But if we consider an arbitrary set far away from the origin, there is no reason for I to have a minimum at a unique point on A, and there is even less reason to have this infimum in the interior of A. In general attains its minimum on a set $\mathcal{D}_A$ depending on A, which can be very messy. Roughly, we will limit ourselves to situations where $\mathcal{D}_A$ is a smooth k-dimensional manifold, which still allows some room for wiggling curves and other rather nasty behaviors. In essence, thinking of A as parametrized by its boundary ∂A, our method consists in applying Laplace's approximation at every point of $\mathcal{D}_A$ along fibers orthogonal to $\mathcal{D}_A$, and integrating these approximations over ∂A, keeping only the leading terms. The difficulty is to obtain a good change of variable formula in order to extract the leading terms, to keep a good control over the Jacobian of the transformation, and to control all the error terms. In doing that, the reader will see that

it amounts to a Laplace method where the parameter λ is now the infinite dimensional quantity ∂A.

Since we allow so much freedom on how the set A can be, we will see that not only should we look at points were I is minimized, but also at points x in A such that $I(x) - I(A)$ stays bounded — this is of course not a rigorous statement — or even is unbounded but not too large compared to some function of $I(A)$.

Approximating an integral of the form $\int_A e^{-I(x)} \mathrm{d}x$ for arbitrary sets A far away from the origin and arbitrary functions I seems a very difficult task. For the applications that we have in mind, we will only concentrate on sets with smooth boundary. We will also require that I be convex. This last assumption may look very restrictive. However, by not putting too many restrictions besides smoothness on ∂A, one can often make a change of variable in order to get back to the case where I is convex. This is actually how we obtained the estimates of the integrals over s_α given at the beginning of this chapter. As the reader can see, $-\log s_\alpha$ is all but a convex function. This idea of changing variables will be systematically illustrated in our examples.

We now outline the content of these notes.

Chapters 2–5 are devoted to the proof of an asymptotic equivalent for integrals of the form $\int_A e^{-I(x)} \mathrm{d}x$ for smooth sets A far away from the origin and convex functions I — plus some other technical restrictions. The goal and culminating point is to prove Theorem 5.1. The main tools come from differential geometry and related formulas in integration. A reader with no interest in the theoretical details can go directly to Theorem 5.1, but will need to read a few definitions in chapters 2–4 in order to fully understand its statement.

In chapter 6, we consider the special case where A is the translate of a fixed set.

A second basic situation is studied in chapter 7, where $A = tA_1$ is obtained by scaling a fixed set A_1 which does not contain the origin, t tends to infinity and I is homogeneous. In this case, we will overlap and somewhat extend the classical Laplace method.

In chapter 8, we study the tail probability of quadratic forms of random vectors. If X is a random vector in $\mathbb{R}^d$ and C is a $d \times d$ matrix, we seek an approximation of $P\{\, \langle CX, X\rangle \geqslant t \,\}$ for large t. We focus on the case where X has a symmetric Weibull type distribution or a Student type distribution. These two cases yield different behavior which we believe to be representative of the so-called sub-Gaussian

and heavy tails distributions. They also illustrate the use of our main approximation result.

The next example, about random linear forms, developed in chapter 9, requires more geometric analysis. The problem is as follows: if X is a random vector on $\mathbb{R}^d$ and M is a subset of $\mathbb{R}^d$, how does $P\{\,\sup_{p\in M}\langle X,p\rangle \geqslant t\,\}$ decay as t tends to infinity? Again, our main theorem provides an answer and we will also focus on symmetric Weibull- and Student-like vectors. These two distributions capture the main features of the applications of our theoretical result to this situation.

The last example, treated in chapter 10, deals with random matrices. Specifically, if X is an $n\times n$ matrix with independent and identically distributed coefficients, we seek for an approximation of $P\{\,\det X \geqslant t\,\}$ for large t. Again, we will deal with Weibull- and Student-like distributions. In the last subsection, we also approximate $P\{\,\|X\|\geqslant t\,\}$ for large t, the tail distribution of the norm of the random matrix. This last example yields some interesting geometry on sets of matrices, and turns to be an application of the results on random linear forms obtained in chapter 9.

The last chapters address some more applied issues, ranging from applications in statistics to numerical computation.

Chapter 11 presents some applications to statistical analysis of time series. We will mainly obtain tail distribution of the empirical covariances of autoregressive models. This turns out to be an application of the results of chapter 8. We will go as far as obtaining numbers useful for some real applications, doing numerical work.

Chapter 12 deals with the distribution of the suprema of some stochastic processes. It contains some examples of pedagogical interest. It concludes with some calculations related to the supremum of the Brownian bridge and the supremum of an amusing process defined on the boundary of convex sets.

There are two appendices. One deals with classical estimates on tail of Gaussian and Student distributions. In the other one, we prove a technical estimate on the exponential map.

Every chapter ends with some notes, giving information and/or open problems on related material.

When doing rather explicit calculations, we will focus on two specific families of distributions. Recall that the Weibull distribution on the positive half line has cumulative distribution function $1-e^{-x^\alpha}$,

$x \geqslant 0$. Its density is $\alpha x^{\alpha-1} e^{-x^\alpha} \mathrm{I}_{[0,\infty)}(x)$. We will consider a variant, namely the density

$$w_\alpha(x) = K_{w,\alpha} e^{-|x|^\alpha/\alpha}\,, \qquad x \in \mathbb{R}\,,$$

where setting $K_{w,\alpha} = \alpha^{1-(1/\alpha)}/\big(2\Gamma(1/\alpha)\big)$ ensures that w_α integrates to 1 on the real line. We call this distribution symmetric Weibull-like. This is the first specific family of distributions that we will use.

To introduce the second family, recall that the Student distribution has density proportional to $(1+\alpha^{-1}y^2)^{-(\alpha+1)/2}$. Its cumulative distribution function $\underline{S}_\alpha$ has tail given by the asymptotic equivalent — see Appendix 1 —

$$\underline{S}_\alpha(-x) \sim 1 - \underline{S}_\alpha(x) \sim \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{x^\alpha} \qquad \text{as } x \to \infty\,,$$

where $K_{s,\alpha} = \int_{-\infty}^{\infty} (1+\alpha^{-1}y^2)^{-(\alpha+1)/2}\,\mathrm{d}y$ is the normalizing constant of the density. Accordingly, we say that a density s_α is Student like if the corresponding distribution function $S_\alpha(x) = \int_{-\infty}^{x} s_\alpha(y)\mathrm{d}y$ satisfies

$$S_\alpha(-x) \sim 1 - S_\alpha(x) \sim \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{x^\alpha} \qquad \text{as } x \to \infty\,,$$

for some constant $K_{s,\alpha}$.

Why are we interested in these two specific families? When $\alpha = 2$ the Weibull-like distribution is the standard Gaussian one. Embedding the normal distribution this way will allow us to see how specific the normal is. When looking at product densities, it is only when $\alpha = 2$ that the Weibull-like distribution is invariant under orthogonal transforms; this will create discontinuities at $\alpha = 2$ in some asymptotic approximations.

The Student-like distributions are of interest because they are rather representative of the so-called heavy tail distributions. In particular, they include the symmetric stable distributions with index α in $(0,2)$. But more generally, recall that a cumulative distribution function F is infinitely divisible if its characteristic function has the form

$$\exp \int_{-\infty}^{+\infty} \frac{e^{i\zeta x} - 1 - i\zeta \sin x}{x^2} \mathrm{d}M(x)$$

for some finite measure M — see, e.g., Feller (1970, §XVII). For x positive, define

$$M^+(x) = \int_x^{+\infty} \frac{\mathrm{d}M(y)}{y^2} \qquad \text{and} \qquad M^-(x) = \int_{-\infty}^{-x} \frac{\mathrm{d}M(y)}{y^2}\,.$$

Recall that a function f on $\mathbb{R}$ is said to be regularly varying with index ρ at infinity if, for any positive λ,

$$\lim_{x\to\infty}\frac{f(\lambda x)}{f(x)}=\lambda^\rho\,.$$

It can be proved — see, e.g., Feller (1970, §XVII.4) — that whenever M^+ (resp. M^-) is regularly varying with negative index, then $1-F(x)\sim M^+(x)$ as x tends to infinity (resp. $F(x)\sim M^-(x)$ as x tends to minus infinity). Thus, if

$$M(-\infty,-x)\sim M(x,\infty)\sim\frac{c}{x^{\alpha-2}}\,,$$

with $\alpha>2$, then F is Student-like. More precisely,

$$1-F(x)\sim\frac{c(\alpha-2)}{\alpha x^\alpha}\qquad\text{as } x\to\infty\,;$$

one can take the constant $K_{s,\alpha}$ to be $c(\alpha-2)\alpha^{-(\alpha+1)/2}$.

Other distributions are Student-like. For instance, one may start with a random variable X having a Pareto distribution on (a,∞),

$$P\{\,X\geqslant x\,\}=\frac{1}{(x-a+1)^\alpha}\,\mathrm{I}_{[a,\infty)}(x)\,.$$

We then symmetrize it. Take ϵ to be a random variable independent of X, with

$$P\{\,\epsilon=-1\,\}=P\{\,\epsilon=+1\,\}=1/2\,.$$

Then ϵX has a symmetric distribution. Its tails are given by

$$P\{\,\epsilon X\geqslant x\,\}=P\{\,\epsilon X\leqslant -x\,\}=\frac{1}{2(x-a+1)^\alpha}$$

for x large enough. Thus, ϵX has a Student-like distribution.

Notes

There is an extensive literature on calculating integrals. Concerning the classical numerical analysis, quadratures, Monte-Carlo methods, etc., chapter 4 of the *Numerical Recipes* by Press, Teukolsky, Vetteling and Flannery (1986) is an excellent starting point. It states clearly the problems, some solutions, their advantages and drawbacks, and contains useful references.

The statistical literature is full of more or less ad hoc methods to perform some specific integrations, sometimes related to those we are interested in here. In the last 20 years or so, Markov chain Monte-Carlo and important sampling methods have been blooming, bringing a large body of papers. Unfortunately, I do not know a reference explaining things simply. Perhaps this area is just too active and has not yet matured to a stage where classical textbooks become available.

Concerning the approximation of univariate integrals, Laplace's method, asymptotic expansions and much more, I find Olver (1974) a great book. The classical references may be Murray (1984); De Bruijn (1958) and Bleistein and Handelsman (1975) have been republished by Dover. Both are inexpensive — as any book should be — and are worth owning. Not so well known is Combet (1982). Combet has proofs of the existence of asymptotic expansions derived from very general principles, although these expansions are not too explicit. I tend to believe that work in this direction would bring some practical results, and this is the first — loose — open problem in these notes.

Regularly varying functions are described beautifully in Bingham, Goldie and Teugels (1987).

2. The logarithmic estimate

Throughout this book, we consider a nonnegative function I defined on $\mathbb{R}^d$, convex, such that

$$\lim_{|x|\to\infty} I(x) = +\infty\,.$$

This assumption holds whenever we use a function I, and we will not repeat it every time.

For any subset A of $\mathbb{R}^d$, we consider the infimum of I over A,

$$I(A) = \inf\{\, I(x) : x \in A \,\}\,.$$

When studying an integral of the form $\int_A e^{-I(x)}\mathrm{d}x$ with A far away from the origin, a first natural question is to investigate if it is close to $e^{-I(A)}$, at least in logarithmically. If this is the case, we have an analogue of the logarithmic estimate for exponential integrals

$$\log \int_A \exp\big(-\lambda I(x)\big)\mathrm{d}x \sim \lambda I(A) \qquad \text{as } \lambda \to \infty$$

— A is fixed here! — which holds, e.g., for sets A with smooth boundary.

Our first result in this flavor will be a slightly sharper upper estimate than a purely logarithmic one, but for some very specific sets A, namely, complements of level sets of I. This estimate will be instrumental in the following chapters.

Let us consider the level sets of I,

$$\Gamma_c = \big\{\, x : I(x) \leqslant c \,\big\}\,, \qquad c \geqslant 0\,.$$

The complement Γ_c^{c} of Γ_c in $\mathbb{R}^d$ is the set of all points for which I is strictly larger than c. Define the function

$$L(c) = \int_{\Gamma_c^{\mathrm{c}}} e^{-I(x)}\mathrm{d}x\,, \qquad c \geqslant 0\,.$$

2.1. PROPOSITION. *There exists a constant C_0 such that for any positive c,*

$$L(c) \leqslant C_0 e^{-c}(1+c)^d .$$

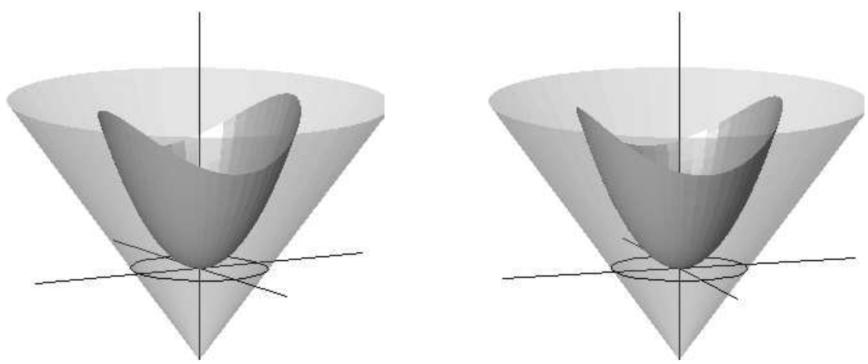
Proof. Use Fubini's theorem and the change of variable $u = v + c$ to obtain

$$\begin{aligned} L(c) &= \int\int \mathrm{I}_{[c,u]}\big(I(x)\big)\,\mathrm{d}x\, e^{-u}\,\mathrm{d}u = e^{-c}\int_{v\geqslant 0} e^{-v}\,|\Gamma_{c+v}\setminus\Gamma_c|\,\mathrm{d}v \\ &\leqslant e^{-c}\int_{v\geqslant 0} e^{-v}\,|\Gamma_{c+v}|\,\mathrm{d}v\,. \end{aligned}$$

Since the function I is convex, nonnegative and tends to infinity with its argument, its graph

$$\big\{\,\big(x, I(x)\big) : x\in\mathbb{R}^d\,\big\} \subset \mathbb{R}^{d+1}$$

is included in a cone with vertex the point $(0,\ldots,0,-1)$ and passing trough a ball centered at 0 in $\mathbb{R}^d\times\{\,0\,\}$.



Thus, $|\Gamma_t|$ is less than the d-dimensional measure of the slice of the cone at height t, i.e.,

$$|\Gamma_t| \leqslant C_1(1+t)^d$$

for some positive constant C_1. Thus,

$$\begin{aligned} L(c) &\leqslant C_1 e^{-c}\int_{v\geqslant 0} e^{-v}(c+v+1)^d\mathrm{d}v \\ &\leqslant 3^d C_1 e^{-c}\int_{v\geqslant 0} e^{-v}(c^d+v^d+1)\mathrm{d}v \end{aligned}$$

and the result follows. ∎

To handle more general sets, it is convenient to introduce the following notation.

NOTATION. *Let F, G be two functions defined on the Borel σ-field of $\mathbb{R}^d$. We write*
(i) $\lim_{A\to\infty} F(A) = 0$ *if and only if for any positive ϵ, there exists c such that for all Borel set A of $\mathbb{R}^d$, the inequality $I(A) \geqslant c$ implies $|F(A)| \leqslant \epsilon$;*
(ii) $\lim_{A\to\infty} F(A) = 0 \Rightarrow \lim_{A\to\infty} G(A) = 0$ *if and only if the following holds:*
$\forall\epsilon$, $\exists\delta$, $\exists c > 0$, $(I(A) \geqslant c$ *and* $|F(A)| \leqslant \delta) \Rightarrow |G(A)| \leqslant \epsilon$.

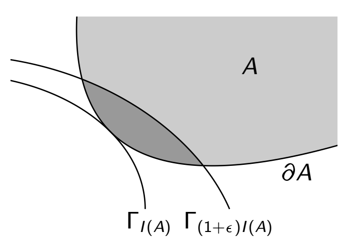
Another way to phrase condition (ii) is to say that whenever $I(A)$ is large enough and $F(A)$ is small enough, then $G(A)$ is small.

Notice that this notion of limit depends on the function I. But if we restrict I to be convex, defined on $\mathbb{R}^d$ and blowing up at infinity, then it does not depend on which such specific function we choose.

The advantage of this notation is that it allows to express approximation properties related to sets moving away from the origin, but under some analytical constraints. We will mainly use (ii). We can similarly define $\liminf_{A\to\infty} F(A)$ and $\limsup_{A\to\infty} F(A)$.

A first example of the use of this notation is to express a condition which ensures that we have a “nice” logarithmic estimate. It asserts that $\log\int_A e^{-I(x)}\mathrm{d}x$ is of order $-I(A)$ provided A is not like a very thin layer attached to $\partial\Gamma_{I(A)}$.

2.2. PROPOSITION. *The following are equivalent:*
(i) $\lim_{A\to\infty} I(A)^{-1}\log\displaystyle\int_A e^{-I(x)}\mathrm{d}x = -1$,
(ii) $\lim_{\epsilon\to 0}\liminf_{A\to\infty} I(A)^{-1}\log|A\cap\Gamma_{(1+\epsilon)I(A)}| = 0$.



REMARK. For any set A,

$$\int_A e^{-I(x)}\mathrm{d}x \leqslant \int_{\Gamma^c_{I(A)}} e^{-I(x)}\mathrm{d}x = L\big(I(A)\big)\,.$$

We then infer from Proposition 2.1 that

$$\limsup_{A\to\infty} I(A)^{-1}\log\int_A e^{-I(x)}\mathrm{d}x \leqslant -1\,.$$

Thus, statement (i) in Proposition 2.2 is really about the limit inferior of the integral as $I(A)$ tends to infinity.

Notice also that the the limit as ϵ tends to 0 in (ii) exists since $\epsilon \mapsto \Gamma_{(1+\epsilon)I(A)}$ is monotone. The limit is at most 0 since $|A\cap\Gamma_{(1+\epsilon)I(A)}| \leqslant |\Gamma_{(1+\epsilon)I(A)}|$ is bounded by a polynomial in $I(A)$. Therefore, statement (ii) is really about the limit inferior as ϵ tends to 0.

Proof of Proposition 2.2. Assume that (ii) holds. For any positive number ϵ,

$$\begin{aligned} I(A)^{-1}\log\int_A e^{-I(x)}\mathrm{d}x &\geqslant I(A)^{-1}\log\int_{A\cap\Gamma_{(1+\epsilon)I(A)}} e^{-I(x)}\mathrm{d}x \\ &\geqslant I(A)^{-1}\log\big(e^{-(1+\epsilon)I(A)}|A\cap\Gamma_{(1+\epsilon)I(A)}|\big) \\ &= -1-\epsilon+I(A)^{-1}\log|A\cap\Gamma_{(1+\epsilon)I(A)}|\,. \end{aligned}$$

Given the above remark, (i) holds.

To prove that (ii) is necessary for (i), let us argue by contradiction. If (ii) does not hold, then there exists a positive β and a sequence of positive numbers ϵ_k converging to 0 as k tends to infinity, such that

$$\liminf_{A\to\infty} I(A)^{-1}\log|A\cap\Gamma_{(1+\epsilon_k)I(A)}| < -\beta$$

for any k large enough. For k large enough, $\epsilon_k < \beta$. Thus, using Proposition 2.1,

$$\begin{aligned} \int_A e^{-I(x)}\,\mathrm{d}x &\leqslant \int_{A\cap\Gamma_{(1+\epsilon_k)I(A)}} e^{-I(x)}\,\mathrm{d}x + \int_{\Gamma^c_{(1+\epsilon_k)I(A)}} e^{-I(x)}\,\mathrm{d}x \\ &\leqslant e^{-I(A)}|A\cap\Gamma_{(1+\epsilon_k)I(A)}| + L\big((1+\epsilon_k)I(A)\big) \\ &\leqslant e^{-I(A)(1+\beta)} + C_0 e^{-I(A)(1+\epsilon_k)}\big(1+(1+\epsilon_k)I(A)\big)^d \\ &\leqslant e^{-I(A)(1+\epsilon_k+o(1))} \qquad \text{as } A\to\infty\,. \end{aligned}$$

Therefore,

$$\limsup_{A\to\infty} I(A)^{-1}\log\int_A e^{-I(x)}\,\mathrm{d}x \leqslant -(1+\epsilon_k) < -1$$

and (i) does not hold. ■

Notes

The idea of the proof of Proposition 2.1, writing e^{-I} as the integral of e^{-u} over $[\,I,\infty)$ and then using Fubini's theorem is all but new; at most it has not been used enough. Lieb and Loss (1997) call a much more general fact the "layer cake representation". I have the recollection of hearing talks using similar tricks and referring to the coarea formula. See Federer (1969) to know all about it, or Morgan (1988) to simply know what it is. Proposition 2.2 grew from Broniatowski and Barbe (200?) which we wrote, I think, in 1996 or 1997. This second chapter has the flavor of large deviation theory and the notation I for the function in the exponential is not a coincidence. More will be said in the notes of chapter 5 and in chapter 7. There are by now a few books on large deviations. Dembo and Zeitouni (1993) and Dupuis and Ellis (1997) are good starting points to the huge literature. From a different perspective, and restricted essentially to the univariate case, Jensen (1995) may be closer to what we are looking for here.

Introducing the set Λ_c suggests that the variations of $-\log f$ are important. It has to do with the following essential remark. The negative exponential function is the only one — up to an asymptotic equivalence — for which integrating on an interval of length of order 1 produces a relative variation of order 1 on the integral. Mathematically, the fact is that

$$\int_{t+s}^{\infty} e^{-x}\,\mathrm{d}x = e^{-s}\int_t^{\infty} e^{-x}\,\mathrm{d}x\,.$$

Hence, in approximating $\int_{t+s}^{\infty} e^{-x}\mathrm{d}x$ by $\int_t^{\infty} e^{-x}\,\mathrm{d}x$, we are making a relative error equals to $e^{-s}-1$. If s is fixed, this relative error stays fixed, event if t moves. If one writes the analogue formula with a power function, one obtains

$$\int_{t+s}^{\infty}\frac{\mathrm{d}x}{\alpha x^{\alpha+1}} = \frac{1}{(t+s)^{\alpha}} = \Big(\frac{t}{t+s}\Big)^{\alpha}\int_t^{\infty}\frac{\mathrm{d}x}{\alpha x^{\alpha+1}}\,.$$

As t tends to infinity, the ratio $\big(t/(t+s)\big)^{\alpha} - 1$ tends to 0. Thus the relative variation of the integral tends to 0 as t tends to infinity. Finally, in the subexponential case, for $\alpha > 1$,

$$\int_{t+s}^{\infty} \alpha x^{\alpha-1} e^{-x^{\alpha}} \,\mathrm{d}x = e^{-(t+s)^{\alpha}+t^{\alpha}} \int_{t}^{\infty} \alpha x^{\alpha-1} e^{-x^{\alpha}} \,\mathrm{d}x \,.$$

The relative variation is now driven by

$$\exp\big(-st^{\alpha-1}(1+o(1))\big) - 1 \,.$$

If s is positive (resp. negative), it tends to -1 (resp. infinity) as t tends to infinity.

In other words, the exponential scale that we use on the variations of f ensures that the variation of $f(x)$ in this scale are of the same order of magnitude as the variations of x in space.

3. The basic bounds

In this chapter, our goal is to obtain lower and upper bounds for the integral $\int_A e^{-I(x)} \mathrm{d}x$. In order to get useful estimates, we need to decompose A into small pieces on which the integral can be almost calculated with a closed formula. These pieces are given by the geometry of the graph of the function I, and therefore we will first devote some time in introducing a few useful quantities related to this graph.

3.1. The normal flow and the normal foliation.

Recall that we assume

> I is strictly convex on $\mathbb{R}^d$, nonnegative, twice differentiable, with $\lim_{x\to\infty} I(x) = +\infty$.

Up to adding a constant to I, which amounts to multiplying the integral $\int_A e^{-I(x)} \mathrm{d}x$ by a constant, and up to translating A by a fixed amount, we can assume that

$$I(0) = 0\,.$$

Let $\mathrm{D}I$ denote the differential — or gradient — of I and $\mathrm{D}^2 I$ its Hessian. The strict convexity assumption implies that $\mathrm{D}^2 I$ is a symmetric definite positive matrix.

Recall that we use

$$\Gamma_c = \left\{\, x \in \mathbb{R}^d : I(x) \leqslant c \,\right\} = I^{-1}([\,0, c\,])$$

to denote the level set of I. We define the level lines — or level hypersurfaces — of I as the points on which I is constant, that is

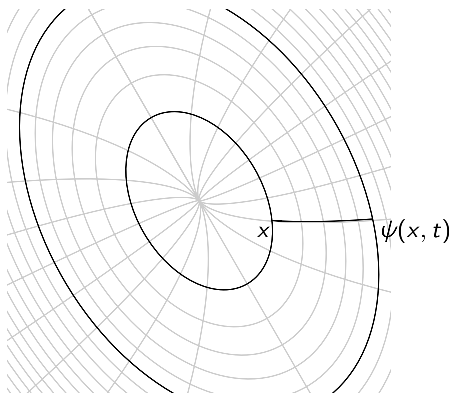
$$\Lambda_c = \left\{\, x \in \mathbb{R}^d : I(x) = c \,\right\} = I^{-1}(\{\, c \,\})\,.$$

Since I is a convex function, Γ_c is a convex set. Moreover, I being defined on $\mathbb{R}^d$, we write

$$\mathbb{R}^d = \bigcup_{c \geqslant 0} \Lambda_c\,.$$

Let x be a nonzero vector in $\mathbb{R}^d$ — equivalently, $I(x)$ is nonzero. Set $c = I(x)$. The gradient $\mathrm{D}I(x)$ is an outward normal to Λ_c at x and defines a vector field on $\mathbb{R}^d \setminus \{0\}$. For t nonnegative, we write $\psi(x,t)$ the integral curve of the vector field $\mathrm{D}I$, such that $\psi(x,0) = x$ and $I\bigl(\psi(x,t)\bigr) = I(x) + t$. Sometime it will be convenient to use the notation $\psi_t(x)$ or $\psi_x(t)$ instead of $\psi(x,t)$.

DEFINITION. *The flow ψ_t is called the normal flow at time t.*



x, the flow $\psi(x,s)$ for s in $[0,t]$,
the level sets $\Lambda_{I(x)}$ and $\Lambda_{I(x)+t}$ hilighted.

It is also convenient to introduce the outward normal unit vector to Λ_c at x,

$$N(x) = \frac{\mathrm{D}I(x)}{|\mathrm{D}I(x)|}\,.$$

3.1.1. LEMMA. *The function $t \mapsto \psi(x,t)$ is a solution of the differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi(x,t) = \frac{\mathrm{D}I\bigl(\psi(x,t)\bigr)}{\bigl|\mathrm{D}I\bigl(\psi(x,t)\bigr)\bigr|^2}\,,$$

with initial condition $\psi(x,0) = x$.

Proof. By definition of the normal flow,

$$1 = \frac{\mathrm{d}}{\mathrm{d}t}I\bigl(\psi(x,t)\bigr) = \Bigl\langle \mathrm{D}I\bigl(\psi(x,t)\bigr), \frac{\mathrm{d}}{\mathrm{d}t}\psi(x,t)\Bigr\rangle\,.$$

The result follows since $\frac{\mathrm{d}}{\mathrm{d}t}\psi(x,t)$ must be collinear to $\mathrm{D}I\bigl(\psi(x,t)\bigr)$. ∎

The next trivial fact is due to the convexity of I. It asserts that the norm of the gradient $|\mathrm{D}I|$ can only increase along the normal flow.

3.1.2. LEMMA. *The function $t \mapsto \bigl|\mathrm{D}I\bigl(\psi(x,t)\bigr)\bigr|$ is nondecreasing.*

Proof. Since $\mathrm{D}^2 I$ is positive definite,

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl|\mathrm{D}I\bigl(\psi(x,t)\bigr)\bigr|^2 = 2\bigl\langle \mathrm{D}^2 I\bigl(\psi(x,t)\bigr)N\bigl(\psi(x,t)\bigr), N\bigl(\psi(x,t)\bigr)\bigr\rangle \geqslant 0\,. \qquad \blacksquare$$

The monotonicity of $|\mathrm{D}I|$ along the normal flow yields a Lipschitz property of the map $t \mapsto \psi(x,t)$.

3.1.3. COROLLARY. *For any x in $\mathbb{R}^d \setminus \{\,0\,\}$ and nonnegative real number t,*

$$|\psi(x,t) - \psi(x,0)| = |\psi(x,t) - x| \leqslant \frac{t}{|\mathrm{D}I(x)|}\,.$$

Proof. The result follows from the fundamental theorem of calculus, Lemmas 3.1.1 and 3.1.2, since

$$\begin{aligned} |\psi(x,t) - \psi(x,0)| \leqslant \int_0^t \Bigl|\frac{\mathrm{d}}{\mathrm{d}s}\psi(x,s)\Bigr|\mathrm{d}s &= \int_0^t \frac{\mathrm{d}s}{\bigl|\mathrm{D}I\bigl(\psi(x,s)\bigr)\bigr|} \\ &\leqslant \frac{t}{|\mathrm{D}I(x)|}\,. \end{aligned} \qquad \blacksquare$$

As a consequence, two points on the same integral curve and nearby level lines of I cannot be too far apart in Euclidean distance. This is expressed in the following result

3.1.4. COROLLARY. *Let M be a positive number. If y is in the forward orbit of x under the flow ψ_t, i.e.,*

$$y \in \{\, \psi_t(x) : t \geqslant 0 \,\}\,;$$

and if $I(y) - I(x)$ is in $[\,0, M\,]$, then

$$|y - x| \leqslant \frac{M}{|\mathrm{D}I(x)|}\,.$$

Proof. Since $y=\psi(x,t)$ for some t,

$$M \geqslant I(y)-I(x)=I\big(\psi(x,t)\big)-I\big(\psi(x,0)\big)=t\,.$$

Hence, Corollary 3.1.3 implies

$$|y-x|\leqslant \frac{t}{|\mathrm{D}I(x)|}\leqslant \frac{M}{|\mathrm{D}I(x)|}$$

∎

So far, we have just collected some trivial informations on the normal flow. Since I is twice differentiable, the level lines Λ_c are submanifolds — hypersurfaces — of $\mathbb{R}^d$. The function ψ_t maps Λ_c onto Λ_{c+t}. For p in Λ_c, we can consider the tangent space $T_p\Lambda_c$ of Λ_c. The differential of ψ_t at some point p, which we denote either by $\mathrm{D}\psi_t(p)$ or $\psi_{t*}(p)$, maps $T_p\Lambda_{I(p)}$ to $T_{\psi_t(p)}\Lambda_{I(p)+t}$. We will need some estimates on this differential, and we first calculate its derivative with respect to t.

3.1.5. LEMMA. *The following holds*

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_{t*}(p)=\Big[\mathrm{Id}-2N\otimes N\big(\psi_t(p)\big)\Big]\frac{\mathrm{D}^2I}{|\mathrm{D}I|^2}\big(\psi_t(p)\big)\psi_{t*}(p)\,.$$

Proof. Lemma 3.1.1 yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_{t*}=\Big(\frac{\mathrm{d}}{\mathrm{d}t}\psi_t\Big)_*=\Big(\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\circ\psi_t\Big)_*=\Big(\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\Big)_*(\psi_t)\circ\psi_{t*}$$

by the chain rule. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_{t*}=\frac{\mathrm{D}^2I}{|\mathrm{D}I|^2}(\psi_t)\circ\psi_{t*}-2\frac{\mathrm{D}I}{|\mathrm{D}I|}\otimes\frac{\mathrm{D}I}{|\mathrm{D}I|}\frac{\mathrm{D}^2I}{|\mathrm{D}I|^2}(\psi_t)\circ\psi_{t*}\,.$$

This is the result since $N=\mathrm{D}I/|\mathrm{D}I|$. ∎

Since I is convex, the level sets Γ_c are convex, and one can see by a drawing that ψ_{t*} must be a map which expands distances — see the picture before Lemma 3.1.1. This is expressed in the next result. Before stating it, notice that for p in Λ_c, the tangent space $T_p\Lambda_c$ inherit of the Hilbert space structure of $\mathbb{R}^d$. Thus, it can be identified with its dual $(T_p\Lambda_c)^*$. Since ψ_{t*} maps $T_p\Lambda_c$ to $T_{\psi_t(p)}\Lambda_{c+t}$, its transpose ψ_{t*}^{T}

maps $(T_{\psi_{t*}(p)}\Lambda_{c+t})^* \equiv T_{\psi_t(p)}\Lambda_{c+t}$ to $T_p\Lambda_c$. Consequently, $\psi_{t*}^{\mathrm{T}}\psi_{t*}$ is a linear map acting on $T_p\Lambda_c$.

3.1.6. LEMMA. *For any nonnegative t, $\det(\psi_{t*}^{\mathrm{T}}\psi_{t*}) \geqslant 1$.*

Proof. Since ψ_0 is the identity function, we have $\det(\psi_{0*}) = \det(\mathrm{Id}) = 1$. It is then enough to prove that for every h in $T_p\Lambda_c$, with $c = I(p)$, the map

$$t \mapsto \bigl|\psi_{t*}(p)h\bigr|^2 = \bigl\langle h\,,\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)h\bigr\rangle$$

is nondecreasing.

Writing $q = \psi_t(p)$, we infer from Lemma 3.1.5 that

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl|\psi_{t*}(p)h\bigr|^2 = 2\Bigl\langle\bigl(\mathrm{Id} - 2N\otimes N(q)\bigr)\frac{\mathrm{D}^2 I}{|\mathrm{D}I|}(q)\psi_{t*}(p)h\,,\psi_{t*}(p)h\Bigr\rangle\,.$$

Since $\psi_{t*}h$ belongs to $T_q\Lambda_{c+t}$ it is orthogonal to $N(q)$. Therefore,

$$N(q)^{\mathrm{T}}\psi_{t*}(p)h = \bigl\langle N(q)\,,\psi_{t*}(p)h\bigr\rangle = 0\,.$$

Since $\mathrm{D}^2 I$ is nonnegative, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl|\psi_{t*}(p)h\bigr|^2 = 2\Bigl\langle\frac{\mathrm{D}^2 I}{|\mathrm{D}I|}(q)\psi_{t*}(p)h\,,\psi_{t*}h\Bigr\rangle \geqslant 0\,.$$

Consequently, the given map is indeed nondecreasing. ∎

It follows from Lemma 3.1.6 that ψ_{t*} is invertible. Since $t \mapsto \psi_t$ is a semigroup, ψ_{t*} cannot expand too fast as t increases.

3.1.7. LEMMA. *For any p in $\mathbb{R}^d\setminus\{\,0\,\}$ and any nonnegative s, t,*

$$\|\psi_{t+s,*}(p)\| \leqslant \|\psi_{t*}(p)\|\exp\biggl(\int_0^s \frac{\bigl\|\mathrm{D}^2 I\bigl(\psi_{t+u}(p)\bigr)\bigr\|}{\bigl|\mathrm{D}I\bigl(\psi_{t+u}(p)\bigr)\bigr|^2}\,\mathrm{d}u\biggr)\,.$$

Proof. Notice that $\mathrm{Id} - 2N\otimes N$ is an inversion. Its operator norm is 1. Using Lemma 3.1.5, writing $q = \psi_t(p)$ and using the triangle inequality for the increments of the function $t \mapsto \bigl\|\psi_{t*}(p)\bigr\|$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl\|\psi_{t*}(p)\bigr\| \leqslant \Bigl\|\frac{\mathrm{d}}{\mathrm{d}t}\psi_{t*}(p)\Bigr\| \leqslant \frac{\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2}(q)\,\|\psi_{t*}(p)\|\,.$$

Integrate this differential inequality between t and s to obtain the result. ∎

In the same spirit as in the previous lemma, a little more work gives us a good control on the growth of $\det(\psi_{t*}^{\mathrm{T}}\psi_{t*})$, improving upon Lemma 3.1.6.

3.1.8. LEMMA. *For any nonzero p and any nonnegative t,*

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\log\det\bigl(\psi_{t*}^{\mathrm{T}}\psi_{t*}(p)\bigr)\\&=\frac{2}{\bigl|\mathrm{D}I\bigl(\psi_t(p)\bigr)\bigr|^2}\mathrm{tr}\Bigl[\mathrm{Proj}_{T_{\psi_t(p)}\Lambda_{I(p)+t}}\mathrm{D}^2I\bigl(\psi_t(p)\bigr)\Big|_{T_{\psi_t(p)}\Lambda_{I(p)+t}}\Bigr]\,.\end{aligned}$$

In particular, for any nonnegative s,

$$\begin{aligned}&\det\bigl(\psi_{t+s,*}^{\mathrm{T}}\psi_{t+s,*}(p)\bigr)\\&\qquad\leqslant\det\bigl(\psi_{t*}^{\mathrm{T}}\psi_{t*}(p)\bigr)\exp\left(2\int_0^s\frac{\mathrm{tr}\bigl[\mathrm{D}^2I\bigl(\psi_{t+u}(p)\bigr)\bigr]}{\bigl|\mathrm{D}I\bigl(\psi_{t+u}(p)\bigr)\bigr|^2}\mathrm{d}u\right),\end{aligned}$$

and the function $s\mapsto\det\bigl(\psi_{t+s,*}^{\mathrm{T}}\psi_{t+s,*}(p)\bigr)$ is nondecreasing.

Proof. Let p be a nonzero vector. Consider a local chart $p(\cdot):U\subset\mathbb{R}^{d-1}\mapsto\Lambda_{I(p)}$ around p, such that $p(0)=p$ and the vectors $\partial_i=\frac{\partial}{\partial u_i}p(0)$ form an orthonormal basis of $T_p\Lambda_{I(p)}$. Lemma 3.1.1 yields

$$\begin{aligned}\frac{\partial}{\partial t}\bigl(\psi_{t*}(p)\partial_i\bigr)&=\frac{\partial^2}{\partial u_i\partial t}\psi\bigl(p(u_1,\ldots,u_{d-1}),t\bigr)\Big|_{u=0}\\&=\frac{\partial}{\partial u_i}\Bigl(\frac{N}{|\mathrm{D}I|}\bigl(\psi(p,t)\bigr)\Bigr)\Big|_{u=0}.\end{aligned}$$

Since $\psi_{t*}\partial_j$ belongs to $T_{\psi_t(p)}\Lambda_{I(p)+t}$, it is orthogonal to $N\bigl(\psi_t(p)\bigr)$ for all j. Moreover, $\mathrm{d}N$ is self-adjoint when acting on the tangent space of the level curves $\Lambda_{I(p)+t}$ — a known fact when studying the second fundamental form of immersions of hypersurfaces in $\mathbb{R}^d$; see, e.g., Do Carmo (1992) — we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl\langle\psi_{t*}(p)\partial_i\,,\psi_{t*}\partial_j\bigr\rangle=\frac{2}{\bigl|\mathrm{D}I\bigl(\psi_t(p)\bigr)\bigr|}\Bigl\langle\mathrm{d}N\bigl(\psi_t(p)\bigr)\psi_{t*}(p)\partial_i\,,\psi_{t*}(p)\partial_j\Bigr\rangle\,.$$

Using that $\mathrm{Ddet}(x) = \det(x)\mathrm{tr}(x^{-1}\cdot)$ — a proof of this fact is in Lemma 10.0.1 — and that ψ_{t*} is invertible,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\det\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)\\
&=\det\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)\mathrm{tr}\Bigl[\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)\Bigr]\\
&=2\,\frac{\det\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)}{\bigl|\mathrm{D}I\bigl(\psi_{t*}(p)\bigr)\bigr|}\,\mathrm{tr}\Bigl[\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)^{-1}\psi_{t*}^{\mathrm{T}}(p)\mathrm{d}N\bigl(\psi_t\ (p)\bigr)\psi_{t*}(p)\Bigr]\\
&=2\,\frac{\det\bigl(\psi_{t*}^{\mathrm{T}}(p)\psi_{t*}(p)\bigr)}{\bigl|\mathrm{D}I\bigl(\psi_{t*}(p)\bigr)\bigr|}\,\mathrm{tr}\Bigl[\mathrm{Proj}_{T_{\psi_t(p)}\Lambda_{I(p)+t}}\mathrm{d}N\bigl(\psi_t(p)\bigr)\Bigr|_{T_{\psi_t(p)}\Lambda_{I(p)+t}}\Bigr]\,.
\end{aligned}$$

This is the first assertion in the Lemma, since the restriction of $\mathrm{d}N$ to the tangent space coincides with $\mathrm{D}^2I/|\mathrm{D}I|$. Moreover, since D^2I is positive, we have

$$\mathrm{tr}\Bigl[\mathrm{Proj}_{T_{\psi_t(p)}\Lambda_{I(p)+t}}\mathrm{d}N\bigl(\psi_t(p)\bigr)\Bigr|_{T_{\psi_t(p)}\Lambda_{I(p)+t}}\Bigr]\leqslant\frac{\mathrm{tr}\bigl[\mathrm{D}^2I\bigl(\psi_t(p)\bigr)\bigr]}{\bigl|\mathrm{D}I\bigl(\psi_t(p)\bigr)\bigr|}\,,$$

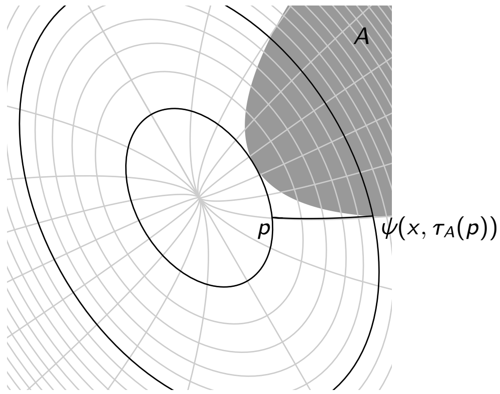
and the second statement follows from integration.

The third statement follows from the first one. Indeed, D^2I is nonnegative, and so $\frac{\mathrm{d}}{\mathrm{d}t}\log\det\bigl(\psi_{t*}^{\mathrm{T}}\psi_{t*}(p)\bigr)\geqslant 0$. ■

We now have all the informations we need on the normal flow. Its integral curves define a foliation of $\mathbb{R}^d\setminus\{\,0\,\}$. We will obtain an expression for $\int_A e^{-I(x)}\mathrm{d}x$ by integrating first on $\Lambda_{I(A)}$, and then along the leaves. The virtue of the normal flow is that computation of the Jacobian in the change of variables is easy, because the normal vector field and the level sets of the function I are orthogonal at *every* point of $\mathbb{R}^d$. Before rewriting the integral, we need to parameterize the boundary ∂A in the flow coordinate system. Specifically, for any point p in $\mathbb{R}^d$, define

$$\tau_A(p)=\inf\bigl\{\,t\geqslant 0:\psi_t(p)\in A\,\bigr\}\,,$$

with the convention that $\inf\emptyset=+\infty$. In the sequel, we will always read $e^{-\infty}$ as 0.



p, the flow $\psi(p,s)$ for s in $[0,\ \tau_A(p)]$,
the level sets $\Lambda_{I(A)}$ and $\Lambda_{I(A)+\tau_A(p)}$ hilighted.

It is convenient to agree on the following convention: Except if specified otherwise, we will equip submanifolds of $\mathbb{R}^d$ with the inner product of $\mathbb{R}^d$ on their tangent spaces. This defines their Riemannian measure completely — not only up to a multiplicative constant.

NOTATION. *Whenever a set M is a submanifold of $\mathbb{R}^d$, we write $\mathcal{M}_M$ for its Riemannian measure.*

We can now rewrite our integral.

3.1.9. PROPOSITION. *If $I(A)$ is positive, the following equality holds,*

$$\int_A e^{-I(x)}\mathrm{d}x = e^{-I(A)} \int_{p\in\Lambda_{I(A)}} e^{-\tau_A(p)} \int_{s\geqslant 0} e^{-s} \frac{\mathrm{I}_A\bigl(\psi_{\tau_A(p)+s}(p)\bigr)}{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)+s}(p)\bigr)\bigr|}$$
$$\bigl|\det\bigl(\psi^{\mathrm{T}}_{\tau_A(p)+s,*}\psi_{\tau_A(p)+s,*}(p)\bigr)\bigr|^{1/2}\mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\,.$$

Proof. Use Fubini's theorem to first obtain

$$\begin{aligned}\int_A e^{-I(x)}\mathrm{d}x &= \int\int \mathrm{I}_A(x)\mathrm{I}_{[I(x),\infty)}(c)e^{-c}\,\mathrm{d}c\,\mathrm{d}x\\ &= e^{-I(A)}\int e^{-c}|A\cap\Gamma_{I(A)+c}|\,\mathrm{d}c\,. \end{aligned}\tag{3.1.1}$$

This brings the leading exponential term out of the integral.

Let p be in $\Lambda_{I(A)}$, and consider a local chart

$$p(\cdot) : U \subset \mathbb{R}^{d-1} \to \Lambda_{I(A)}$$

around p. If x is in the image of $p(U)$ through the normal flow, we can parameterize it as $x = \psi\bigl(p(u_1,\ldots,u_{d-1}),s\bigr)$. Setting $\partial_i = \frac{\partial}{\partial u_i}p$, the Jacobian of the change of variable $x \leftrightarrow (u_1,\ldots,u_{d-1},s)$ is

$$\begin{aligned}
J &= \Bigl|\det\Bigl(\psi_{s*}(p)\partial_1,\ldots,\psi_{s*}(p)\partial_{d-1},\frac{\partial}{\partial s}\psi(p,s)\Bigr)\Bigr| \\
&= \det\left|\begin{pmatrix} \bigl\langle \psi_{s*}(p)\partial_i\,,\psi_{s*}(p)\partial_j\bigr\rangle_{1\leqslant i,j\leqslant d-1} & 0 \\ 0 & \Bigl|\dfrac{\partial}{\partial s}\psi\bigl(p(u),s\bigr)\Bigr|^2 \end{pmatrix}\right|^{1/2} \\
&= \Bigl|\frac{\mathrm{d}}{\mathrm{d}s}\psi_s\Bigr|\bigl(\det(\psi_{s*}^{\mathrm{T}}\psi_{s*})\bigr)^{1/2}\det\bigl(\langle\partial_i,\partial_j\rangle_{1\leqslant i,j\leqslant d-1}\bigr)^{1/2},
\end{aligned}$$

since $\bigl\langle\psi_{s*}(p)\partial_i\,,\frac{\mathrm{d}}{\mathrm{d}s}\psi(p,s)\bigr\rangle = 0$, for $i=1,\ldots,d-1$.

In term of the new parameterization and gluing charts using a partition of unity, we obtain

$$\begin{aligned}
|A\cap\Gamma_{I(A)+c}| &= \int \mathrm{I}_{A\cap\Gamma_{I(A)+c}}(x)\mathrm{d}x \\
&= \int_{0\leqslant s\leqslant c}\int_{\Lambda_{I(A)}} \mathrm{I}_A\bigl(\psi(p,s)\bigr)\Bigl|\frac{\mathrm{d}}{\mathrm{d}s}\psi(p,s)\Bigr|\times \\
&\qquad\qquad \det(\psi_{s*}^{\mathrm{T}}\psi_{s*})^{1/2}\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\,\mathrm{d}s\,.
\end{aligned}$$

Notice that $\mathrm{I}_A\bigl(\psi(p,s)\bigr)$ vanishes if $s\leqslant\tau_A(p)$. Using (3.1.1) and Fubini's theorem, the expression of $\mathrm{d}\psi(p,s)/\mathrm{d}s$ in Lemma 3.1.1 gives the result. ∎

Just as we introduced $\tau_A(p)$, the first ent<u>r</u>ance time of a point in A through the flow, define

$$\chi_A^L(p) = \sup\bigl\{\, s : \psi_{\tau_A(p)+s}(p)\in A\,\bigr\}$$

the <u>l</u>ast time of e<u>x</u>it of A starting the clock time at $\tau_A(p)$, and

$$\chi_A^F(p) = \inf\bigl\{\, s>0 : \psi_{\tau_A(p)+s}(p)\notin A\,\bigr\}\,,$$

the <u>f</u>irst time of e<u>x</u>it of A starting the time at $\tau_A(p)$. If A has "holes", χ_A^F may be strictly less than χ_A^L. Both $\chi_A^L(p)$ and $\chi_A^F(p)$ may be infinite.

Proposition 3.1.9 and our lemmas on the normal flow yield the following basic bounds, which will turn out to be surprisingly sharp and useful.

3.1.10. THEOREM. *For any Borel set A of $\mathbb{R}^d$ with $I(A)$ positive,*

$$\begin{aligned}\int_A e^{-I(x)}\mathrm{d}x \leqslant e^{-I(A)} \int_{p\in\Lambda_{I(A)}} & e^{-\tau_A(p)} \frac{\det\bigl(\psi^{\mathrm{T}}_{\tau_A(p)*}(p)\psi_{\tau_A(p)*}(p)\bigr)^{1/2}}{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)}(p)\bigr)\bigr|}\times \\ & \int_{0\leqslant s\leqslant \chi_A^L(p)} \exp\Bigl[\int_{0\leqslant t\leqslant s}\Bigl(d\frac{\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2}\bigl(\psi_{\tau_A(p)+t}(p)\bigr)-1\Bigr)\mathrm{d}t\Bigr] \\ & \mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\end{aligned}$$

and

$$\begin{aligned}\int_A e^{-I(x)}\mathrm{d}x \geqslant e^{-I(A)} \int_{\substack{p\in\Lambda_{I(A)}\\ \tau_A(p)<\infty}} & e^{-\tau_A(p)} \frac{\det(\psi^{\mathrm{T}}_{\tau_A(p)*}(p)\psi_{\tau_A(p)*}(p))^{1/2}}{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)}(p)\bigr)\bigr|}\times \\ & \int_{0\leqslant s\leqslant \chi_A^F(p)} \exp\Bigl[\int_{0\leqslant t\leqslant s}\Bigl(\frac{-\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2}\bigl(\psi_{\tau_A(p)+t}(p)\bigr)-1\Bigr)\mathrm{d}t\Bigr] \\ & \mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\,.\end{aligned}$$

(Be aware that the letter d in front of $\|\mathrm{D}^2 I\|/|\mathrm{D}I|$ in the exponential in the upper bound refers to the dimension d of $\mathbb{R}^d$, and not to some differentiation!)

Proof. The upper bound follows from Proposition 3.1.9, and the following observations. Clearly, $\mathrm{I}_A\bigl(\psi_{\tau_A(p)+s}(p)\bigr)$ is at most 1, and vanishes whenever $s \geqslant \chi_A^L(p)$. Lemma 3.1.8 implies

$$\begin{aligned}&\det\bigl(\psi^{\mathrm{T}}_{\tau_A(p)+s,*}\psi_{\tau_A(p)+s,*}\bigr) \\ &\quad\leqslant \det\bigl(\psi^{\mathrm{T}}_{\tau_A(p),*}\psi_{\tau_A(p),*}\bigr)\exp\Bigl(2d\int_{0\leqslant t\leqslant s}\frac{\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2}(\psi_{\tau_A(p)+t})\mathrm{d}t\Bigr)\,,\end{aligned}$$

and Lemma 3.1.2 gives

$$\frac{1}{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)+s}(p)\bigr)\bigr|}\leqslant\frac{1}{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)}(p)\bigr)\bigr|}\,.$$

To prove the lower bound, notice first that $\mathrm{I}_A\big(\psi_{\tau_A(p)+s}(p)\big)=1$ for all s in $\big[\,0,\chi_A^F(p)\big)$. Using Lemma 3.1.8 and Proposition 3.1.9,

$$\int_A e^{-I(x)}\mathrm{d}x \geqslant e^{-I(A)}\int_{p\in\Lambda_{I(A)}} e^{-\tau_A(p)}\frac{\det\big(\psi^{\mathrm{T}}_{\tau_A(p)*}\psi_{\tau_A(p)}(p)\big)}{\big|\mathrm{D}I\big(\psi_{\tau_A(p)}(p)\big)\big|}\times$$
$$\int_{0\leqslant s\leqslant\chi_A^F(p)}\exp\Big(-s-\log\big|\mathrm{D}I\big(\psi_{\tau_A(p)+s}(p)\big)\big|+\log\big|\mathrm{D}I\big(\psi_{\tau_A(p)}(p)\big)\big|\Big)$$
$$\mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\,.$$

Using Lemma 3.1.1, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\log\Big|\mathrm{D}I\big(\psi(p,s)\big)\Big| = \Big\langle\frac{\mathrm{D}^2I\cdot N}{|\mathrm{D}I|^2},N\Big\rangle\big(\psi(p,s)\big)\,.$$

Therefore,

$$\log\big|\mathrm{D}I\big(\psi_{\tau_A(p)}(p)\big)\big| - \log\big|\mathrm{D}I\big(\psi_{\tau_A(p)+s}(p)\big)\big|$$
$$\geqslant\int_{0\leqslant t\leqslant s}-\frac{\|\mathrm{D}^2I\|}{|\mathrm{D}I|^2}\big(\psi_{\tau_A(p)+t}(p)\big)\mathrm{d}t\,,$$

and this brings the lower estimate. ∎

REMARK. The gap between the upper and the lower bounds comes essentially from the term $\|\mathrm{D}^2I\|/|\mathrm{D}I|^2$ in the exponential. One should expect this ratio to be small for large arguments and in interesting situations. For instance, when $d=1$ and $I(x)=|x|^\alpha$, we have

$$\frac{|\mathrm{D}^2I(x)|}{|\mathrm{D}I(x)|^2}=\frac{\alpha-1}{\alpha}\frac{1}{|x|^\alpha}\,.$$

In the same vein, if $\tau_A(p)$ is not too large, we should be able to replace $\psi_{\tau_A(p)}(p)$ by $\psi_0(p)=p$, while if $\tau_A(p)$ is large, the term $e^{-\tau_A(p)}$ will make the contribution of those p's negligible. Therefore, we hope to obtain the approximation

$$\begin{aligned}\int_A e^{-I(x)}\mathrm{d}x\\ &\approx e^{-I(A)}\int_{p\in\Lambda_{I(A)}}\frac{e^{-\tau_A(p)}}{\big|\mathrm{D}I(p)\big|}\int_{s\geqslant 0}\exp\Big(\int_0^s -1\mathrm{d}t\Big)\mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\\ &= e^{-I(A)}\int_{p\in\Lambda_{I(A)}}\frac{e^{-\tau_A(p)}}{\big|\mathrm{D}I(p)\big|}\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\,, \qquad (3.1.2)\end{aligned}$$

which is a quite manageable expression. To prove that such approximation is valid requires some ideas on the order of magnitude of the final expression in (3.1.2), and the next section is devoted to the study of this term.

3.2. Base manifolds and their orthogonal leaves.

Following the remark concluding the previous subsection, we want to rewrite the integral as $e^{-I(A)}$ times

$$\int_{\Lambda_{I(A)}} \frac{e^{-\tau_A}}{|\mathrm{D}I|} \mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}\,, \tag{3.2.1}$$

so that we can isolate the leading terms. We first need to recall some notation and introduce some definitions.

For p in Λ_c, we denote by $\exp_p(\cdot)$ the exponential map at p in the manifold Λ_c. That is, if u belongs to $T_p\Lambda_c$, the value of $\exp_p(u)$ is the point on Λ_c at a distance $|u|$ to p on the geodesic starting at p in the direction u — see for instance Do Carmo (1992) or Chavel (1996). The exponential map is always defined for a small value of the argument in the tangent plane.

If M is a submanifold of $\Lambda_{I(A)}$ and p is a point in M, the tangent space $T_p\Lambda_{I(A)}$ splits as $T_pM \oplus (T_pM)^\perp$ where we denote by $(T_pM)^\perp$ the orthocomplement of T_pM in $T_p\Lambda_{I(A)}$.

DEFINITION. *The projection of a set $A \subset \mathbb{R}^d$ on $\Lambda_{I(A)}$ — through the flow ψ — is the set of all points p in $\Lambda_{I(A)}$ such that $\psi_t(p)$ belongs to A for some positive t. A submanifold $\mathcal{D}_A \subset \Lambda_{I(A)}$ is called a base manifold — for A — if the projection of A through ψ is contained in $\bigcup_{p\in\mathcal{D}_A} \exp_p\big((T_p\mathcal{D}_A)^\perp\big)$.*

Often, we will loosely speak of the projection of A on $\Lambda_{I(A)}$, forgetting to add that it is through the flow ψ.

In what follows, if $\mathcal{D}_A$ is a base manifold, we denote by $k = k(\mathcal{D}_A)$ its dimension. In order to replace the integration over $\Lambda_{I(A)}$ by an integration over a base manifold $\mathcal{D}_A$ in (3.2.1), we need to attach orthogonal leaves to $\mathcal{D}_A$. For this construction, we borrow some definitions and notations used to parameterize tubes as in Weyl's (1939) formula.

Consider a base manifold $\mathcal{D}_A$ of A, and define a normal bundle by

$$N_p\mathcal{D}_A = T_p\Lambda_{I(A)} \ominus T_p\mathcal{D}_A\,, \qquad p \in \mathcal{D}_A\,.$$

Let $\mathrm{dist}(\cdot\,,\cdot)$ denote the Riemannian distance on $\Lambda_{I(A)}$. For every u in $N_p\mathcal{D}_A$, consider the radius of injectivity of $\Lambda_{I(A)}$ in the direction u,

$$e_A(p,u) = \sup\big\{\, t \geqslant 0 : \mathrm{dist}\big(\exp_p(tu), p\big) = t \,\big\}\,,$$

the exponential map being that on $\Lambda_{I(A)}$ as before. Define

$$\Omega_{A,p} = \big\{\, (t,u) : u \in N_p\mathcal{D}_A\,,\ |u| = 1\,,\ 0 \leqslant t < e_A(p,u) \,\big\}\,, \qquad p \in \mathcal{D}_A\,,$$

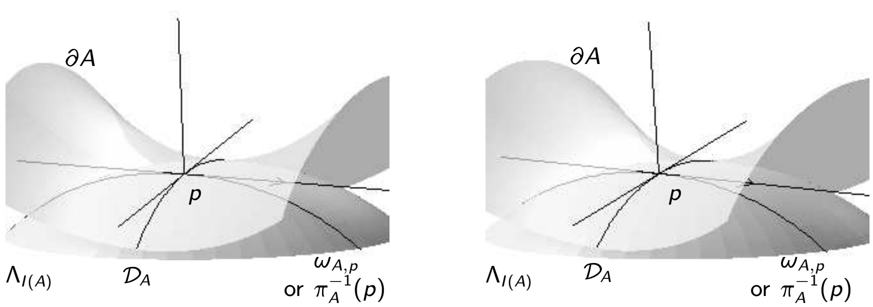
so that $\bigcup_{p\in\mathcal{D}_A} \partial\Omega_{A,p}$ is the set of all focal points of $\mathcal{D}_A$ immersed in $\Lambda_{I(A)}$. From its definition, we infer that $\exp_p(\Omega_{A,p})$ coincide with the set

$$\omega_{A,p} = \{\, q \in \Lambda_{I(A)} : \text{ there exists a unique minimizing geodesic through } q \text{ which meets } \mathcal{D}_A \text{ orthogonally at } p \,\}.$$

Set

$$\omega_A = \bigcup_{p\in\mathcal{D}_A} \omega_{A,p}\,.$$

On ω_A, we can define a projection π_A onto $\mathcal{D}_A$ as follows. Any point q in ω_A can be written in a unique way as $q = \exp_p(u)$ for $(p,u) \in \mathcal{D}_A \times N_p\mathcal{D}_A$. We set $\pi_A(q) = p$. In other words, q is on a unique geodesic starting from $\mathcal{D}_A$ and orthogonal to $\mathcal{D}_A$; the projection of q on $\mathcal{D}_A$ is the starting point of this geodesic in $\mathcal{D}_A$. We call the sets $\{\,\pi_A^{-1}(p) : p \in \mathcal{D}_A\,\}$, the orthogonal leaves to $\mathcal{D}_A$. By construction, $\omega_{A,p}$ is in $\Lambda_{I(A)} = \Lambda_{I(p)}$, and it contains $\pi_A^{-1}(p)$.



In order to rewrite the integral (3.2.1) as an integral over a base manifold and its orthogonal leaves, we need to calculate the Jacobian of the change of variable $q \in \omega_A \leftrightarrow \exp_p(u)$, $p \in \mathcal{D}_A$, $u \in N_p\mathcal{D}_A$. For this purpose, notice that the differential $\pi_{A*}(q)$ maps $T_q\Lambda_{I(A)}$

onto $T_{\pi(q)}\mathcal{D}_A$. Since π_A is constant on the orthogonal leaves, the orthocomplement of $\ker \pi_{A*}(q)$ is a vector space of dimension k. Let $b_1(q), \ldots, b_k(q)$ be an orthonormal basis of this orthocomplement, and define the Jacobian

$$J\pi_A(q) = \Bigl|\det\Bigl(\bigl(\pi_{A*}(q) \wedge \ldots \wedge \pi_{A*}(q)\bigr)(b_1, \ldots, b_k)\Bigr)\Bigr|$$

— when $k = 0$, read $J\pi_A(q) = 1$. Federer's (1959) co-area formula — see the appendix of Howard (1993) for a simple proof in the smooth case we are using here — yields

$$\int_{\omega_A} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|} \mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p) = \int_{u\geqslant 0} \int_{p\in\mathcal{D}_A} \int_{q\in\pi_A^{-1}(p)} e^{-u} \frac{\mathrm{I}_{(-\infty,u]}\bigl(\tau_A(q)\bigr)}{\bigl|\mathrm{D}I(q)\bigr|}$$
$$(J\pi_A)^{-1}(q)\mathrm{d}\mathcal{M}_{\pi_A^{-1}(p)}(q)\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\mathrm{d}u\,. \qquad (3.2.2)$$

We can now express the Riemannian measure $\mathcal{M}_{\pi_A^{-1}(p)}$ over the leaf $\pi_A^{-1}(p)$ in normal coordinates. For this aim, following Chavel (1993, chapter 3), let $\mathcal{T}_{p,t,v}$ denote the parallel transport along the geodesic from p to $\exp_p(tv)$, and furthermore, let R_q denote the Riemannian curvature tensor of the leaf $\pi_A^{-1} \circ \pi_A(q)$ at the point q. Denote $\gamma_{p,v}(t) = \exp_p(tv)$ a point of the geodesic starting from p in the direction v with constant velocity $|v|$. Then, consider the matrix

$$\mathcal{R}_v(t) = \mathcal{T}_{p,t,v}^{-1} \circ R_{\exp_p(tv)}\bigl(\gamma'_{p,v}(t), \cdot\bigr)\gamma'_{p,v}(t) \circ \mathcal{T}_{p,t,v}$$

defined on $T_p\pi_A^{-1}(p)$. We define a matrix-valued function $\mathcal{A}_p(t,v)$ as solving the differential equation — in the set of matrices over the $(d-k-2)$-dimensional vector space $T_p\pi_A^{-1}(p) \ominus v\mathbb{R} = N_p\mathcal{D}_A \ominus v\mathbb{R}$ —

$$\frac{\partial^2}{\partial t^2}\mathcal{A}(t,v) + \mathcal{R}_v(t)\mathcal{A}(t,v) = 0\,,$$

subject to the boundary condition

$$\mathcal{A}(0,v) = 0\,, \qquad \frac{\partial}{\partial t}\mathcal{A}(t,v)\Big|_{t=0} = \mathrm{Id}\,.$$

The boundary conditions imply

$$\det\mathcal{A}(t,v) \sim \det(t\,\mathrm{Id}_{\mathbb{R}^{d-k-2}}) \sim t^{d-k-2} \qquad \text{as } t \to 0$$

uniformly in v in the unit sphere $S_{T_p\pi_A^{-1}(p)}(0,1)$ of $N_p\mathcal{D}_A$. The expression of the Riemannian measure over a leaf, $\mathcal{M}_{\pi_A^{-1}(p)}$, in normal coordinates yields

$$\begin{aligned}\int_{\omega_A} & \frac{e^{-\tau_A(q)}}{|\mathrm{D}I(q)|}\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(q)\\ &= \int_{u\geqslant 0} e^{-u}\int_{p\in\mathcal{D}_A}\int_{v\in S_{T_p\pi_A^{-1}(p)}}\int_{t\in[0,e_A(p,v)]}\frac{\mathrm{I}_{[0,u]}\Bigl(\tau_A\bigl(\exp_p(tv)\bigr)\Bigr)}{\bigl|\mathrm{D}I\bigl(\exp_p(tv)\bigr)\bigr|}\\ &\qquad\frac{1}{\bigl|J\pi_A\bigl(\exp_p(tv)\bigr)\bigr|}\det\bigl(\mathcal{A}(t,v)\bigr)\,\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}u\,, \quad (3.2.3)\end{aligned}$$

where μ_p is the Riemannian measure over the unit sphere of $T_p\pi_A^{-1}(p)$, centered at the origin.

This last expression looks quite complicated. However, we are almost done, and the intuition goes as follows. Roughly, we want to choose $\mathcal{D}_A$ as $\Lambda_{I(A)}\cap\partial A$, so that I is minimal in A over $\mathcal{D}_A$ — but we will actually need to have a little bit of freedom for some applications and make a slightly more subtle choice. Due to the term e^{-u}, let us concentrate on the range $u = O(1)$. The assumption $|\mathrm{D}I(p)|$ tends to infinity with $|p|$ will imply that in good situations $\tau_A\bigl(\exp_p(tv)\bigr)$ grows very fast as a function of t, since $\exp(tv)$ is transverse to $\mathcal{D}_A$ where τ_A is minimal. Hence, if the indicator function of $\tau_A\bigl(\exp_p(tv)\bigr)\leqslant u = O(1)$ is not zero, we must have t small. But for small t's and $p\in\mathcal{D}_A$,

$$(J\pi_A)^{-1}\bigl(\exp_p(tv)\bigr)\approx(J\pi_A)^{-1}(p)=1$$

since $\exp_p(tv)\approx\exp_p(0)=p$ and $\pi_A(p)=p$; furthermore, as we mentioned earlier, for small t's

$$\det\bigl(\mathcal{A}(t,v)\bigr)\approx t^{d-k-2}$$

and again, since $\exp_p(tv)\approx p$,

$$\bigl|\mathrm{D}I\bigl(\exp_p(tv)\bigr)\bigr|\approx\bigl|\mathrm{D}I(p)\bigr|\,.$$

Thus, we should expect the right hand side of (3.2.3) to be approximately

$$\begin{aligned}&\int_{u\geqslant 0} e^{-u}\int_{p\in\mathcal{D}_A}\frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|}\int_{v\in\mathcal{S}_{T_p\pi_A^{-1}(p)}(0,1)}\int_{t\in[0,e_A(p,v)]}\\ &\mathrm{I}_{(-\infty,u]}\Bigl(\tau_A\bigl(\exp_p(tv)\bigr)-\tau_A(p)\Bigr)t^{d-k-2}\,\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}u\,. \quad (3.2.4)\end{aligned}$$

Assuming that ∂A is smooth, for a unit vector v in $T_p\pi_A^{-1}(p)$, the function $t \mapsto \tau_A\big(\exp(tv)\big) - \tau_A(p)$, is minimum for $t = 0$. This function should be approximately quadratic near 0. The integration in t and v then gives the volume of an ellipsoid, which is related to the curvatures of $\Lambda_{I(A)}$ and ∂A near p. We will also need to prove that in good situations (3.2.1) is equivalent to (3.2.2), in which we restricted the integration to ω_A.

At this stage it should be noticed that pinching the curvature R yields differential inequalities for $\mathcal{A}$ and therefore bounds for (3.2.3). We shall not pursue this line, but the reader may notice that there are situations where R is easy to calculate — for instance if $I(x) = |x|^2$ as in the Gaussian case, then $R = \mathrm{Id}/|x|$ — or to pinch. Thus, more precise estimates could be obtained, and even a control of the error terms. This could be useful in some applications, but it is not clear that it is worth investigating in a general setting.

Notes

This chapter builds upon the classical theory of surfaces and integration on manifolds. If you don't know any differential geometry, don't give up! It took me a long time to find a good starting point, that is a book that I could read and understand. I found it when I was visiting the Université Laval at Quebec! Buy Do Carmo's (1976, 1992) two books, and start reading the one on curves and surfaces. If you are as bad learner as I am, do what I did, that is, all the exercises. Once you read about two dimensional surfaces and understand that curvature is a geometric name for a second order Taylor formula, you will have enough intuition to digest the abstract Riemannian manifolds — which really copy the classical theory of surfaces in $\mathbb{R}^3$. After reading Do Carmo's books, I found Chavel (1996) and some parts of Spivak (1970) most valuable. Some colleagues liked Morgan's (1992) book very much as a starting point, others McCleary's (1994).

The change of variable comes from that in Weyl's (1939) tube formula. Weyl's paper is very nice to read, and a good part does not require much knowledge of differential geometry.

Notice that the formula derived in Proposition 3.1.9 does not rely upon convexity of I. It could be of some use in other places.

It is certainly possible to extend the estimates obtained in this chapters, the previous ones and the next ones, to some integral over noncompact manifolds. In this case, one should replace the Lebesgue

measure by the Riemannian one. I do not know if this could be of any use.

4. Analyzing the leading term for some smooth sets

In this chapter we analyze further the quantities involved in the integral (3.2.3). We will assume that

$$\partial A \text{ and } \Lambda_{I(A)} \text{ are smooth — i.e., } \mathrm{C}^2 \text{ — submanifolds of } \mathbb{R}^d, \tag{4.0.1}$$

and furthermore that

the base manifold $\mathcal{D}_A$ is a smooth — i.e., C^2 — submanifold of $\mathbb{R}^d$, possibly with boundary. (4.0.2)

Notice that so far we have a lot of freedom to choose the base manifold, so that (4.0.2) is not much of a restriction. Borrowing from the theory of large deviations, it is convenient to introduce the following notion.

DEFINITION. *A base manifold $\mathcal{D}_A$ for A is called a dominating manifold (for A) if, for all p in $\mathcal{D}_A$ and all unit vectors v in $T_p\pi_A^{-1}(p)$, the function*

$$t \in \big(-e_A(p,-v), e_A(p,v)\big) \mapsto \tau_A\big(\exp_p(tv)\big) \in [0,\infty)$$

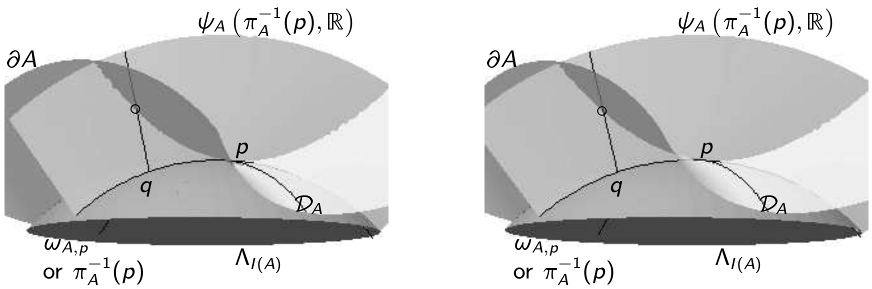
has a local finite minimum at $t=0$. A point in a dominating manifold is called a dominating point.

Thus, if $\mathcal{D}_A$ is a dominating manifold, the set ∂A is pulling away from $\Lambda_{I(p)}$ along the lift through the normal flow of the geodesics orthogonal to $\mathcal{D}_A$. Notice the important fact that if a point p is in a dominating manifold $\mathcal{D}_A$, it may not be in ∂A, i.e., $\tau_A(p)$ may still be positive. This is an essential difference from the large deviation theory, or analogously to what would be considered in Laplace's method. This distinction will turn out to be crucial for some applications — see sections 8.2, 8.3 or 10.2. The downside of allowing this extra freedom is that the main result has a slightly more involved statement. But it is worth the extra power provided. Another feature is that only the points for which $\tau_A\big(\exp_p(tv)\big)$ is finite matter.

4.1. Quadratic approximation of τ_A near a dominating manifold.

Under the assumptions (4.0.1)–(4.0.2), if $\mathcal{D}_A$ is a dominating manifold and p is one of its points, the function $t \mapsto \tau_A\bigl(\exp_p(tv)\bigr)$ is minimal at 0 for all unit vectors v tangent to $\pi_A^{-1}(p)$ at p. It admits a quadratic approximation. To obtain it, let us denote by $\Pi_{\Lambda_c,q}$ (resp. $\Pi_{\partial A,q}$) the second fundamental form of Λ_c (resp. ∂A) at $q \in \Lambda_c$ (resp. $q \in \partial A$). Those are defined using the unit outward normal vector N to Λ_c. At points of $\Lambda_{I(A)} \cap \partial A$, this unit normal N is also a unit normal for ∂A, and so ∂A is oriented by an extension of N. The submanifold $\pi_A^{-1}(p) \subset \Lambda_{I(A)} \subset \mathbb{R}^d$ admits a second fundamental form also associated with the normal field N. For q belonging to $\pi_A^{-1}(p)$, we denote it by $\Pi^{\pi}_{\Lambda_{I(A)},q}$. It is nothing but the restriction of $\Pi_{\Lambda_{I(A)},q}$ to $T_q\pi_A^{-1}(p)$. The leaf $\pi_A^{-1}(p)$ can be lifted to ∂A through the normal flow in considering the set

$$\psi_A\bigl(\pi_A^{-1}(p)\bigr) = \bigl\{\, \psi\bigl(q, \tau_A(q)\bigr)\, ,\, q \in \pi_A^{-1}(p) \,\bigr\}\, .$$



On this picture, the line leaving the point q is the normal flow $\psi(q,t)$ for $t \geqslant 0$. It crosses the boundary ∂A at the circled point $\psi(q, \tau_A(q))$. When q moves on the geodesic $\pi_A^{-1}(p)$, the crossing point describes $\psi_A(\pi^{-1}(p))$.

This lifted leaf admits a second fundamental form relative to the extension of N on ∂A. At any point p in $\mathcal{D}_A$ for which $\tau_A(p)$ vanishes, this second fundamental form is the restriction of $\Pi_{\partial A,p}$ to the tangent space of the lifted manifold, which coincides with the tangent space $T_p\pi_A^{-1}(p) = N_p\mathcal{D}_A$. We denote by $\Pi^{\pi}_{\partial A,p}$ this restriction. The difference of the fundamental forms, $\Pi^{\pi}_{\Lambda_{I(A)},p} - \Pi^{\pi}_{\partial A,p}$ can be interpreted as follows. Assume we live on the leaf $\pi_A^{-1}(p) \subset \Lambda_{I(A)}$, and look at ∂A along the "vertical" direction given by the normal flow. From $\pi_A^{-1}(p)$, the boundary ∂A pulls away as we move away from

$p \in \pi_A^{-1}(p) \cap \partial A$. The difference of the fundamental forms is a measure of the curvature of ∂A viewed from $\pi_A^{-1}(p)$, with vertical distance measured in Euclidean distance along the normal flow. However, as far as the integration goes, the right measure of vertical distance is on the increments of the function I. At an infinitesimal level, the increment is $1/|\mathrm{D}I(p)|$ near p — see Lemma 3.1.1. Thus, in the geometry of the level sets of I, the bending of ∂A away from $\Lambda_{I(A)}$ is measured by $|\mathrm{D}I(p)|(\Pi^\pi_{\Lambda_{I(A)}} - \Pi^\pi_{\partial A,p})$.

Equipped with this interpretation, the following result is very natural.

4.1.1. PROPOSITION. *Let $p(s)$ be a curve on $\pi_A^{-1}(p)$, such that $p(0) = p$ belongs to $\partial A \cap \Lambda_{I(A)}$. Under (4.0.1)–(4.0.2),*

$$\tau_A\big(p(s)\big) = \frac{s^2}{2}|\mathrm{D}I(p)|\big\langle(\Pi^\pi_{\Lambda_{I(A)},p} - \Pi^\pi_{\partial A,p})p'(0)\,,p'(0)\big\rangle + o(s^2)$$

as s tends to zero.

Proof. We will make use of the following elementary fact. If $q(s)$ is a curve on a surface M and ν is a unit normal vector field on M, then $\big\langle\nu(q(s)), q'(s)\big\rangle = 0$. Differentiating with respect to s,

$$\langle \mathrm{d}\nu(q)q', q'\rangle + \langle\nu(q), q''\rangle = 0\,.$$

Since p belongs to $\partial A \cap \Lambda_{I(A)}$, the function $q \in \Lambda_{I(A)} \mapsto \tau_A(q) \in \mathbb{R}$ is minimal at p. Hence, $\mathrm{D}\tau_A(p)$ is collinear to $\mathrm{D}I(p)$. Differentiating at $s = 0$ the relation $\tau_A(q) - s = \tau_A\big(\psi(q,s)\big)$ yields $\mathrm{D}\tau_A(p) = -\mathrm{D}I(p)$.

Next, let $p(s)$ be a curve on $\Lambda_{I(A)}$ such that $p(0) = p$. Let $t = p'(0)$ be its tangent vector at p, and set $v = p''(0)$. Define

$$\theta_1 = -\langle \mathrm{D}I(p)\,, v\rangle + \langle \mathrm{D}^2\tau_A(p)t\,, t\rangle\,.$$

A Taylor expansion yields $\tau_A\big(p(s)\big) = s^2\theta_1/2 + o(s^2)$. To prove Proposition 4.1.1, we just need to find an expression for θ_1. Denote

$$g(s) = \psi\big[p(s), \tau_A\big(p(s)\big)\big] = \psi\big(p(s), s^2\theta_1/2 + o(s^2)\big)\,,$$

the curve $p(s)$ lifted to ∂A through the normal flow. A Taylor expansion and Lemma 3.1.1 gives

$$g(s) = p + st + \frac{s^2}{2}\Big(v + \theta_1\frac{\mathrm{D}I}{|\mathrm{D}I|^2}(p)\Big) + o(s^2)\,.$$

In particular, the curves $\psi\big[p(s),\tau_A\big(p(s)\big)\big]$ and $p(s)$ have the same tangent vector t at $s=0$. Writing ν for a unit normal vector field extending $N(p)$ on ∂A, using the Weingarten map, and the elementary fact at the beginning of this proof,

$$\begin{aligned}\langle \Pi_{\partial A,p}t\,,t\rangle &= \langle \mathrm{d}\nu(p)g'(0)\,,t\rangle = -\langle \nu(0)\,,g''(0)\rangle\\ &= \Big\langle N(p), v+\theta_1\frac{N(p)}{|\mathrm{D}I(p)|}\Big\rangle\,.\end{aligned}$$

This gives an expression for θ_1 involving

$$\langle N(p)\,,v\rangle = \Big\langle N(p)\,,\frac{\mathrm{d}^2}{\mathrm{d}s^2}p(s)\big|_{s=0}\Big\rangle = -\langle \mathrm{d}N(p)t\,,t\rangle = \langle \Pi_{\Lambda_{I(A)},p}t\,,t\rangle$$

— using the elementary fact again — and ultimately the result. ∎

4.2. Approximation for det$\mathcal{A}(t,v)$.

In this section we obtain some bounds on the determinant of the matrix $\mathcal{A}(t,v)$ involved in formula (3.2.3). They will be instrumental in approximating (3.2.3) further. These bounds are given by the inequalities of Gunther (1960) and Bishop (1977) — see, e.g., Chavel (1993, pp. 118-121). They involve the curvature tensor of the surface $\pi_A^{-1}(p)\subset\Lambda_{I(p)}$ and so we will first compute that of the level set Λ_c.

In order to avoid any ambiguity, recall that if E is a subspace of $\mathbb{R}^d$, the compression of a $d\times d$ matrix M to E is the linear operator from E into E obtained in restricting M to E and then projecting the image of M into E. Thus, writing Proj_E for the projection from $\mathbb{R}^d$ to E and $|_E$ for the restriction to E, the compression of M to E is $\mathrm{Proj}_E\circ M|_E$.

4.2.1. LEMMA. *Let $S(p)$ be the compression of $\mathrm{D}^2I(p)/\big|\mathrm{D}I(p)\big|$ to $T_p\Lambda_{I(p)}$. The curvature tensor of the surface $\Lambda_{I(p)}$ at p is given by*

$$\begin{aligned}R : x,y,z\in T_p\Lambda_{I(p)} &\mapsto R(x,y)z\\ &= \langle S(p)x,y\rangle S(p)y-\langle S(p)y,z\rangle S(p)x\in T_p\Lambda_{I(p)}\,.\end{aligned}$$

Proof. Let $\mathrm{Proj}_{T_p\Lambda_{I(p)}}$ denote the orthogonal projection from $\mathbb{R}^d$ onto $T_p\Lambda_{I(p)}$. Furthermore, we write ∇ for the standard Riemannian connection on $\mathbb{R}^d$. For any x in $T_p\Lambda_{I(p)}$,

$$-\mathrm{Proj}_{T_p\Lambda_{I(p)}}\nabla_x N(p) = \frac{1}{|\mathrm{D}I(p)|}\mathrm{Proj}_{T_p\Lambda_{I(p)}}\nabla_x\mathrm{D}I(p) = S(p)x\,.$$

It follows that the second fundamental form of $\Lambda_{I(p)} \subset \mathbb{R}^d$ is the bilinear map

$$B : x, y \in T_p\Lambda_{I(p)} \mapsto B(x,y) = \langle S(p)x, y\rangle N(p) \in \mathbb{R}^d \ominus T_p\Lambda_{I(p)} .$$

The result follows from the immersion of $\Lambda_{I(p)}$ in the space $\mathbb{R}^d$ of null curvature, and say, Theorem 2.2 in Chavel (1993). ∎

We now obtain an expression for the sectional and the Ricci curvature of $\Lambda_{I(p)}$.

4.2.2. LEMMA *The sectional curvature of the surface $\Lambda_{I(p)}$ at p is*

$$K : x, y \in T_p\Lambda_{I(p)} \mapsto K(x,y) = \frac{\langle S(p)x,x\rangle\langle S(p)y,y\rangle - \langle S(p)x,y\rangle^2}{|x|^2|y|^2 - \langle x,y\rangle^2} \in \mathbb{R} .$$

Its Ricci curvature is

$$\mathrm{Ricc} : x, y \in T_p\Lambda_{I(A)} \mapsto \mathrm{Ricc}(x,y) = \sum_{1\leqslant j\leqslant d-1} \big(\langle S(p)x,y\rangle\langle S(p)e_j,e_j\rangle - \langle S(p)x,e_j\rangle\langle S(p)e_j,y\rangle\big) \in \mathbb{R} ,$$

where $e_1, \ldots, e_{d-1}$ is any orthogonal basis of $T_p\Lambda_{I(p)}$.

Proof. The expression of the sectional curvature follows from Theorem 2.2 in Chavel (1993) say, and the calculation of the Weingarten map $\mathrm{Proj}_{T_p\Lambda_{I(p)}}\nabla_x N$ in the proof of Lemma 4.2.1. The expression for the Ricci curvature follows from Lemma 4.2.1. ∎

We can now pinch — i.e., bound above and bellow — the curvatures, and for this purpose, let us denote $\lambda_{\min}\big(\mathrm{D}^2 I(p)\big)$ and $\lambda_{\max}\big(\mathrm{D}^2 I(p)\big)$ the smallest and the largest eigenvalues of the symmetric definite positive $d \times d$ matrix $\mathrm{D}^2 I(p)$.

4.2.3. PROPOSITION. *For any x in $T_p\Lambda_{I(p)}$,*

$$\mathrm{Ricc}(x,x) \geqslant (d-2)\frac{\lambda_{\min}\big(\mathrm{D}^2 I(p)\big)}{|\mathrm{D}I(p)|^2}|x|^2 .$$

If x, y are in $T_p\Lambda_{I(p)}$ and are orthogonal — i.e., $\langle x,y\rangle = 0$ — then

$$K(x,y) \leqslant \frac{\lambda_{\max}\big(\mathrm{D}^2 I(p)\big)^2}{|\mathrm{D}I(p)|^2}|x||y| .$$

Proof. Since $\mathrm{D}^2 I$ is a symmetric definite positive matrix, so is $S(p)$. Let $0 \leqslant s_1 \leqslant \cdots \leqslant s_{d-1}$ be the eigenvalues and $e_1, \ldots, e_{d-1}$ be the corresponding orthonormal basis of eigenvectors of $S(p)$. Let us denote by $x = \sum_{1\leqslant i\leqslant d-1} x_i e_i$ and $y = \sum_{1\leqslant j\leqslant d-1} y_i e_i$ two vectors in $T_p\Lambda_{I(p)}$. We then have

$$\begin{aligned}\mathrm{Ricc}(x,x) &= \sum_{1\leqslant j\leqslant d-1}\Big(\sum_{1\leqslant i\leqslant d-1} x_i^2 s_i s_j - x_j^2 s_j^2\Big)\\ &= \sum_{1\leqslant i\leqslant d-1} x_i^2 s_i \Big(\mathrm{tr}\big(S(p)\big) - s_i\Big)\,.\end{aligned}$$

If $s_{d-1} \geqslant \mathrm{tr}\big(S(p)\big)/2$, the function $s \in [\,s_1, s_{d-1}\,] \mapsto s\big(\mathrm{tr}S(p) - s\big)$ is increasing on $[\,s_1, \mathrm{tr}S(p)/2\,]$, decreasing on $[\,\mathrm{tr}S(p)/2, s_{d-1}\,]$.

Otherwise, if $s_{d-1} \leqslant \mathrm{tr}S(p)/2$, this function is increasing. Thus, in any case, for any $i = 1, \ldots, d-1$,

$$\begin{aligned}s_i\big(\mathrm{tr}S(p) - s_i\big) &\geqslant s_1\big(\mathrm{tr}S(p) - s_1\big) \wedge s_{d-1}\big(\mathrm{tr}S(p) - s_{d-1}\big)\\ &\geqslant (d-2)s_1^2\,.\end{aligned}$$

Since $S(p)$ is a compression of $\mathrm{D}^2 I(p)/|\mathrm{D}I(p)|$, the smallest eigenvalue of $\mathrm{D}^2 I(p)/|\mathrm{D}I(p)|$ is larger than s_1. The bound on the Ricci curvature follows.

To obtain an upper bound for the sectional curvature, notice that if x and y are of unit norm and orthogonal, then

$$\begin{aligned}K(x,y) = \sum_{1\leqslant i\leqslant d-1} s_i x_i^2 &\sum_{1\leqslant i\leqslant d-1} s_i y_i^2 - \Big(\sum_{1\leqslant i\leqslant d-1} x_i y_i s_i\Big)^2 \leqslant s_{d-1}s_{d-1}\\ &\leqslant \lambda_{\max}\big(\mathrm{D}^2 I(p)\big)^2/|\mathrm{D}I(p)|^2\,,\end{aligned}$$

where the last inequality comes from the fact that $S(p)$ is a compression of $\mathrm{D}^2 I(p)/\mathrm{D}I(p)$. ∎

Let us denote by $\mathrm{inj}_{\pi_A^{-1}(p)}(p)$ the radius of injectivity of p in the manifold $\pi_A^{-1}(p)$. Since $\pi_A^{-1}(p)$ is built of geodesics of $\Lambda_{I(p)}$, we have

$$\mathrm{inj}_{\pi_A^{-1}(p)}(p) = \inf\big\{\, c(p,u) \wedge e_A(p,u) : u \in T_p\pi_A^{-1}(p)\,,\ |u| = 1\,\big\}\,,$$

where $c(p,u)$ is the Riemannian distance from p to its cut point on $\Lambda_{I(p)}$ along the geodesic leaving p in the direction u — notice that this

property on $\mathrm{inj}_{\pi_A^{-1}(p)}(p)$ is very specific to the way we constructed the leaves, and to the fact that we only consider the point p.

From the previous lemmas, we can deduce some bounds on $\det\mathcal{A}(t,v)$, using some classical volume comparison theorems.

4.2.4. PROPOSITION. *(i) (Bishop-Gunther) Assume that for any unit vector v in $T_p\pi_A^{-1}(p)$, the Ricci curvature along the geodesic $\exp_p(tv)$ in $\pi_A^{-1}(p)$ is nonnegative. Then, for all t in $\big[\,0, \mathrm{inj}_{\pi_A^{-1}(p)}(p)\big)$, the inequality $\det\mathcal{A}(t,v) \leqslant t^{d-k-2}$ holds.*
(ii) For any positive t, let

$$K_{\max}(p,t) = \sup\Big\{\, \frac{\lambda_{\max}\big(\mathrm{D}^2 I(q)\big)^2}{|\mathrm{D}I(q)|^2} : q \in \pi_A^{-1}(p)\,,\, d(p,q) \leqslant t \,\Big\}\,.$$

Then, for any positive t_0 and any t in $\Big[\,0, \dfrac{\pi}{\sqrt{K_{\max}(p,t_0)}} \wedge e_A(p,v)\Big)$,

$$\det\mathcal{A}(t,v) \geqslant \left[\frac{\sin\big(\sqrt{K_{\max}(p,t_0)}t\big)}{\sqrt{K_{\max}(p,t_0)\big)}}\right]^{d-k-2}.$$

Proof. (i) is Bishop's (1977) or Günther's (1960) theorem. Assertion (ii) follows from Bishop's (1977) theorem — see, e.g., Chavel, 1996, pp.118-120 — provided we have the proper upper bound on the Ricci curvature along the geodesic $\exp_p(tv)$ in $\pi_A^{-1}(p)$. For x,y in $T_q\pi_A^{-1}(p)$, denote $K_{\pi_A^{-1}(p),q}(x,y)$ the sectional curvature at q of the manifold $\pi_A^{-1}(p)$. Then (ii) follows from the bound

$$K_{\pi_A^{-1}(p),q}(x,y) \leqslant K_{\Lambda_{I(p)},q}(x,y) \leqslant \frac{\lambda_{\max}\big(\mathrm{D}^2 I(q)\big)^2}{\|\mathrm{D}I(q)\|^2}\,.$$

To prove this inequality, extend y to a vector field in the tangent bundle $T\pi_A^{-1}(p)$. The second fundamental form of the immersion $\pi_A^{-1}(p) \subset \Lambda_{I(p)}$ is given by $B(x,y) = \mathrm{Proj}_{T_q\pi_A^{-1}(p)}(\nabla_x y)$ where ∇ is the connection on $\Lambda_{I(p)}$. In particular, if x is parallel — which is the case for $x = \frac{\mathrm{d}}{\mathrm{d}t}\exp_p(tv)$ — then $B(x,x) = 0$. Hence — see, e.g., Theorem 2.2 in Chavel, 1996 — if x and y are of unit norm,

$$K_{\pi_A^{-1}(p),q}(x,y) = K_{\Lambda_{I(p)},q}(x,y) - \big|B(x,y)\big|^2 \leqslant K_{\Lambda_{I(q)},q}\,.$$

We conclude in using Proposition 4.2.3. ∎

4.3. What should the result be?

Combining the results of sections 4.1 and 4.2, the heuristic argument at the end of section 3 and (3.1.2), we see that we should expect

$$\int_A e^{-I(x)}\mathrm{d}x \approx e^{-I(A)} \int_{u\geqslant 0} e^{-u} \int_{p\in\mathcal{D}_A} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|} \int_{v\in S_{T_p\pi_A^{-1}(p)}(0,1)} \int_{t\in[0,e_A(p,v)]} \mathrm{I}_{[0,u]}\Big(\frac{1}{2}|\mathrm{D}I(p)|\big\langle G_A(p)tv,tv\big\rangle\Big)t^{d-k-2} \,\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}u\,,$$

where $G_A(p) = \Pi_{\Lambda_{I(A)},p} - \Pi_{\partial A,p}$. Assuming for the time being that everything works as expected, we can complete the calculation of the asymptotic equivalent. Recall that the volume of the unit ball of $\mathbb{R}^n$ is $\omega_n = \pi^{n/2}/\Gamma(\frac{n}{2}+1)$.

4.3.1. PROPOSITION. *For any a,b both positive, the following holds,*

$$\lim_{u_0\to\infty} \int_{u\in[0,u_0]} e^{-au} \int_{p\in\mathcal{D}_A} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|} \int_{v\in S_{T_p\pi_A^{-1}(p)}(0,1)} \int_{t\geqslant 0} \mathrm{I}_{[0,bu]}\Big(\frac{1}{2}|\mathrm{D}I(p)|\langle G_A(p)tv,tv\rangle\Big)t^{d-k-2}\,\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}u$$

$$= \frac{b^{(d-k-1)/2}}{a^{(d-k+1)/2}} 2^{(d-k-1)/2}\Gamma\Big(\frac{d-k+1}{2}\Big)\omega_{d-k-1}\times \int_{p\in\mathcal{D}_A} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|^{(d-k+1)/2}\big(\det G_A(p)\big)^{1/2}}\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,.$$

REMARK. In Proposition 4.3.1, the integral in t is in the range $[0,\infty)$, while it is in the range $[0,e_A(p,v))$ in the integral at the beginning of this section. But this will not make any difference ultimately.

Proof. Notice that $t^{d-k-2}\mathrm{d}t\,\mathrm{d}\mu_p(v)$ is the Lebesgue measure in the tangent space $T_p\pi_A^{-1}(p)\equiv\mathbb{R}^{d-k-1}$. Therefore,

$$\int_{v\in S_{T_p\pi_A^{-1}(p)}(0,1)} \int_{t\geqslant 0} \mathrm{I}_{[0,bu]}\Big(\frac{1}{2}|\mathrm{D}I(p)|\big\langle G_A(p)tv,tv\big\rangle\Big)t^{d-k-2}\mathrm{d}t\,\mathrm{d}\mu_p(v)$$

$$= \Big| \Big\{ x \in \mathbb{R}^{d-k-2} : \big\langle G_A(p)x, x \big\rangle \leqslant \frac{2bu}{|\mathrm{D}I(p)|} \Big\} \Big|_{\mathbb{R}^{d-k-1}}$$

$$= \Big(\frac{2bu}{|\mathrm{D}I(p)|} \Big)^{(d-k-1)/2} \frac{\omega_{d-k-1}}{\big(\det G_A(p)\big)^{1/2}} \,.$$

Hence, the integral for which we want to take the limit as u_0 tends to infinity is

$$\int_{u\in[0,u_0]} e^{-au} \int_{p\in\mathcal{D}_A} \frac{e^{-\tau_A(p)}}{\big|\mathrm{D}I(p)\big|^{(d-k+1)/2}} (2b)^{(d-k-1)/2} \omega_{d-k-1} \times \\ u^{(d-k-1)/2} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)}{\big(\det G_A(p)\big)^{1/2}} \, \mathrm{d}u \,.$$

We perform the integration in u and let u_0 tends to infinity to obtain the term $\Gamma\big((d-k+1)/2\big)$ in the Proposition. ∎

5. The asymptotic formula

We now have everything that we need to derive our asymptotic approximation for $\int_A e^{-I(x)}\mathrm{d}x$ as the set A tends to infinity nicely. The assumptions that we require may look quite bad at first glance. However, the reader will see in the next chapters that they are in fact quite well tailored for applications.

We will assume that for any fixed positive M, there exists some positive number $c_{A,M}$, depending on A and meeting the following requirements.

We first assume that there exists a manifold $\mathcal{D}_A$ such that

$$\begin{array}{l}\mathcal{D}_A \text{ is a dominating manifold for the set } A\cap\Gamma_{I(A)+c_{A,M}}\,,\\ \text{of fixed dimension } k.\end{array} \tag{5.1}$$

Hypothesis, (5.1) contains two key requirements. First, k does not depend on A, restricting the class of sets A that we consider. Second, if $\mathcal{D}_A$ is a dominating manifold, it has to be a base manifold. Therefore, ∂A can pull away from $\mathcal{D}_A$ only in the orthogonal directions. This is a restriction for instance if $\mathcal{D}_A$ is a closed curve with two boundary points; we may want to allow ∂A to pull in the outward tangent directions at the boundary points. However, by breaking A into smaller pieces, the restriction can be overcome in practice.

Let us denote by $\underline{A}_M$ the projection of $A\cap\Gamma_{I(A)+c_{A,M}}$ on $\Lambda_{I(A)}$ through the normal flow ψ, that is

$$\underline{A}_M = \{\, p\in\Lambda_{I(A)} : \tau_A(p)\leqslant c_{A,M}\,\}\,.$$

Under (5.1), it makes sense to assume that

$$\underline{A}_M \subset \bigcup_{p\in\mathcal{D}_A}\omega_{A,p}\,, \tag{5.2}$$

where $\omega_{A,p}$ is defined in section 3.2.

When ∂A is smooth, Proposition 4.1.1 gives a quadratic approximation of τ_A. Of course, such an approximation is local and does not have any kind of uniformity with respect to A. Moreover, looking at (3.2.4), we also would like to have a quadratic approximation for

$\tau_A(q)-\tau_A(p)$ for q near $\psi\bigl(p,\tau_A(p)\bigr)$. How near? Well, we need to cover $\underline{A}_M$. But Proposition 4.1.1 suggests that

$$\frac{1}{2}\Bigl|\mathrm{D}I\Bigl(\psi\bigl(p,\tau_A(p)\bigr)\Bigr)\Bigr|\bigl\langle\bigl(\Pi_{\Lambda_{I(p)+\tau_A(p)}}-\Pi_{\partial A\,,\psi(p,\tau_A(p))}\bigr)(q-p),q-p\bigr\rangle$$

should be a good approximation for $\tau_A(q)-\tau_A(p)$. Since $c_{A,M}$ is expected to be small compare to $I(p)=I(A)$, we should have $\psi\bigl(p,\tau_A(p)\bigr)$ quite close to p, and so maybe the quadratic approximation given in Proposition 4.1.1 is just fine. This is what happens in many interesting examples. But for the time being, there is no other way than to force it, and assume that there exists a linear map $G_A(p)$ on $T_p\Lambda_{I(A)}$ such that

$$\lim_{A\to\infty}\sup_{q\in\underline{A}_M}\left|\frac{\tau_A(q)-\tau_A\bigl(\pi_A(q)\bigr)}{\frac{1}{2}\bigl|\mathrm{D}I\bigl(\pi_A(p)\bigr)\bigr|\bigl\langle G_A\bigl(\pi_A(q)\bigr)\exp^{-1}_{\pi_A(q)}(q)\,,\exp^{-1}_{\pi_A(q)}(q)\bigr\rangle}-1\right| = 0\,. \qquad (5.3)$$

This can be rewritten as

$$\lim_{A\to\infty}\sup_{\substack{v\in T_p\Lambda_{I(A)}\ominus T_p\mathcal{D}_A\\ \tau_A(\exp_p(|v|))\leqslant c_{A,t}}}\left|\frac{\tau_A\bigl(\exp_p(v)\bigr)-\tau_A(p)}{\frac{1}{2}|\mathrm{D}I(p)|\langle G_A(p)v,v\rangle}-1\right|=0\,.$$

Now that we have a dominating manifold and a quadratic approximation, it makes sense to proceed as the heuristic argument at the end of section 3 suggests. Then, we can guess the asymptotic equivalent of the integral in using the result of section 4.3. Define

$$c^*_{A,M}=\inf\left\{\,c: L\bigl(I(A)+c\bigr)\leqslant\frac{1}{M}e^{-I(A)}\int_{\mathcal{D}_A}\frac{e^{-\tau_A}d\mathcal{M}_{\mathcal{D}_A}}{|\mathrm{D}I|^{\frac{d-k+1}{2}}(\det G_A)^{\frac{1}{2}}}\right\}.$$

Since all our asymptotic analysis is driven by the desire to have the integral influenced mainly by the behavior of ∂A near $\mathcal{D}_A$, and to reduce the integral to $A\cap\Gamma_{I(A)+c_{A,M}}$ where things go well, we assume that

$$c_{A,M}\geqslant c^*_{A,M}\,. \qquad (5.4)$$

Any $c_{A,M}$ larger than $c^*_{A,M}$ will do, and the reader will easily see that the larger $c_{A,M}$ is, the more stringent our assumptions are. Hence, $c^*_{A,M}$ is the best choice, but can seldom be calculated exactly. In practice, picking for $c_{A,M}$ an asymptotic equivalent of a multiple of

$c^*_{A,M}$ will do. To fix the ideas, a typical order of magnitude of $c_{A,M}$ is $\log I(A)$ for the applications that we will study.

Though we want to be able to localize the study of the integral to points of A near $\mathcal{D}_A$, we still want A to have some thickness! In particular, we do not want the main contribution in the integral to come from the thinness of A — think for instance of taking $A = \Gamma_{c+\epsilon} \setminus \Gamma_c$ for $\epsilon = \exp(-e^c)$ or even smaller, and looking for asymptotics as c tends to infinity. This can be ruled out by assuming that the first exit time of the normal flow after a time τ_A is large enough, namely, that for all positive M,

$$\lim_{A\to\infty} \inf_{p\in\underline{A}_M} \chi^F_A(p) = +\infty\,. \tag{5.5}$$

Since χ^F_A is less or equal to χ^L_A, assumption (5.5) implies

$$\lim_{A\to\infty} \inf_{p\in\underline{A}_M} \chi^L_A(p) = +\infty\,.$$

In the same spirit, we see that it does not make any difference in the asymptotic analysis if we replace A by $A \cap \Gamma_{I(A)+c_{A,M}}$. Thus, if $c_{A,M}$ stays bounded, the set A is quite small in the geometry of the level sets of I. In this case, we would need to deal with an analogue of Proposition 4.3.1, but keeping u_0 fixed. This would introduce an incomplete gamma function. In order to simplify the result, we assume that $c_{A,M}$ is chosen such that

$$\lim_{A\to\infty} c_{A,M} = +\infty\,. \tag{5.6}$$

This assumption is satisfied in all the applications that follow; but the reader will see that up to changing some constant in the asymptotics, the proof still goes through without it.

Our next condition can be explained by first thinking of two difficult situations. Imagine that around the dominating manifold $\mathcal{D}_A$, the set ∂A is pulling away very slowly from $\Lambda_{I(A)}$. Thus, going in the normal direction to $\mathcal{D}_A$ on $\Lambda_{I(A)}$, we need to take q very far away from $\mathcal{D}_A$ in order to have $\tau_A(q)$ not too close to 0. In such circumstances, we should need an extra rescaling so that in the new scale τ_A grows faster.

Another difficult situation would be to have the level sets Γ_c concentrated along a proper subspace of $\mathbb{R}^d$ — think for instance, if $d = 2$, of some ellipsoid with one axis growing like e^c and the other like c. On a large scale, the problem would be essentially

lower dimensional. Notice however that in such circumstances, the curvature of the level set should be of different orders of magnitude at different points. Near points of high curvature, the normal flow could pull away very slowly in the Euclidean geometry. And so there is not much hope to localize the problem near the boundary of A.

A way to take care of these different situations is to relate the curvature of the level sets with the rate at which ∂A pulls away from $\mathcal{D}_A$ in the normal directions. Define

$$t_{0,M}(p) = \sup\Big\{\, t : \inf_{v\in S_{T_p\pi_A^{-1}(p)}(0,1)} \tau_A\big(\exp_p(tv)\big) \leqslant c_{A,M} \,\Big\}\,.$$

Whenever q in $\pi_A^{-1}(p)$ is at distance $t_{0,M}(p)$ or more from p, it satisfies $\tau_A(q) \geqslant c_{A,M}$. We then assume that for all positive M,

$$\lim_{A\to\infty} \sup_{p\in\mathcal{D}_A} \sqrt{K_{\max}\big(p, t_{0,M}(p)\big)}\, t_{0,M}(p) = 0\,. \tag{5.7}$$

In order to be able to use Proposition 4.2.4.i, we impose that for any p in $\mathcal{D}_A$ and any unit vector v in $T_p\pi_A^{-1}(p)$,

the Ricci curvature along the geodesic $\exp_p(tv)$ in $\pi_A^{-1}(p)\cap \underline{A}_M$ is nonnegative. (5.8)

— I am inclined to believe that convexity of I is enough to guarantee (5.8) but could not prove it.

It is also possible that the set A and the function I are such that $|\mathrm{D}I|$ varies widely in a small neighborhood of some point in the dominating manifold. In such situation, I would increase very fast in some specific directions, and much slower in others. Then, the rescaling needed, even in $\pi_A^{-1}(p)$ — when $\pi_A^{-1}(p)$ is of dimension at least 2 — could not be homogeneous in different directions. This can be ruled out by assuming that for any positive M,

$$\lim_{A\to\infty} \sup\Big\{\, \Big|\frac{|\mathrm{D}I(p)|}{|\mathrm{D}I(q)|} - 1\Big| : p\in\mathcal{D}_A\,,\ q\in \underline{A}_M\cap\pi_A^{-1}(p) \,\Big\} = 0\,. \tag{5.9}$$

In order to proceed along the lines of the final remark in section 3.1, we need that

$$\lim_{A\to\infty} \sup\Big\{\, \sup_{t\geqslant 0} \frac{\big\|\mathrm{D}^2 I\big(\psi_t(p)\big)\big\|}{\big|\mathrm{D}I\big(\psi_t(p)\big)\big|^2} : p\in \underline{A}_M\,,\ \tau_A(p)<\infty \,\Big\} = 0\,. \tag{5.10}$$

This assumption mainly controls the growth of I. A sufficient condition to guarantee (5.10) is of course to have $\|\mathrm{D}^2 I(p)\|/|\mathrm{D}I(p)|$ tends to 0 as $|p|$ tends to infinity, which means essentially that I grows slower than any exponential function.

Actually, we also need a rate of convergence in (5.10), but in a rather weak sense, namely that

$$\lim_{A\to\infty} \sup \left\{ \int_0^{\tau_A(p)} \frac{\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2} \big(\psi_u(p)\big)\,\mathrm{d}u : p \in \underline{A}_M \right\} = 0\,. \tag{5.11}$$

Finally, we need two technical assumptions in order to carry over the intuitive argument at the end of section 3.2, namely that for any positive M,

$$\lim_{A\to\infty} \sup_{q\in\underline{A}_M} \big|J\pi_A(q) - 1\big| = 0\,, \tag{5.12}$$

and that for any fixed positive w, and, as A moves to infinity,

$$\begin{aligned} &\int_{\substack{\mathcal{D}_A \\ \tau_A \geqslant c_{A,M}-w}} \frac{e^{-\tau_A}}{|\mathrm{D}I|^{\frac{d-k+1}{2}} (\det G_A)^{\frac12}}\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A} \\ &\qquad\qquad = o\bigg(\int_{\mathcal{D}_A} \frac{e^{-\tau_A}}{|\mathrm{D}I|^{\frac{d-k+1}{2}} (\det G_A)^{\frac12}}\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A} \bigg)\,. \end{aligned} \tag{5.13}$$

Often $\tau_A(\cdot)$ vanishes on $\mathcal{D}_A$. Then (5.13) is satisfied whenever (5.6) is. In particular, this is always the case when $\mathcal{D}_A$ is a single point or the union of a finite number of points. Similarly, (5.11) and (5.12) always hold if $\mathcal{D}_A$ reduces to a point.

The reader may legitimately be suspicious about these assumptions, and how they can be checked in applications. We will show in nontrivial examples that they are not too difficult to verify. They reduce the problem of approximating the integral to much more manageable small problems, which can be handled by systematic methods. The key point to understand is perhaps that as $c_{A,M}$ is usually much smaller than $I(A)$, the set $\underline{A}_M$ is actually quite close to $\mathcal{D}_A$ in the scale given by I.

We can now state our main result.

5.1. THEOREM. *Assume that $\partial A \cap \Gamma_{I(A)+c_{A,M}}$ is a smooth — twice differentiable — manifold for every positive M. Then, under (5.1)–(5.13),*

$$\int_A e^{-I(x)}\mathrm{d}x \sim e^{-I(A)} (2\pi)^{\frac{d-k-1}{2}} \int_{\mathcal{D}_A} \frac{e^{-\tau_A}}{|\mathrm{D}I|^{\frac{d-k+1}{2}} (\det G_A)^{\frac12}}\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}$$

as A moves to infinity, and with $k = \dim \mathcal{D}_A$.

REMARK. If $\mathcal{D}_A$ is an open subset of $\Lambda_{I(A)}$, then $k = d-1$, and the formula should be read with the determinant of G_A to be 1.

Proof. We obtain an upper and a lower bound for the integral.

Let us start with the upper bound. Let ϵ be a positive number. Assume first that, for M large enough,

$$A = A \cap \Gamma_{I(A)+c_{A,M}} \tag{5.14}$$

provided $I(A)$ is large enough. Thus, for p in $\Gamma_{I(A)}$, either $\tau_A(p)$ is less than $c_{A,M}$ or is infinite.

Recall that ψ_{0*} is the identity since ψ_0 is the identity as well. Consequently, assumption (5.11), Lemmas 3.1.6 and 3.1.8 imply that for $\tau_A(p) \leqslant c_{A,M}$,

$$\begin{aligned} 1 \leqslant \det(\psi^{\mathrm{T}}_{\tau_A(p),*}\psi_{\tau_A(p),*})^{1/2} \\ \leqslant \exp\left(d\int_0^{\tau_A(p)} \frac{\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2}\big(\psi_u(p)\big)\mathrm{d}u\right) \leqslant 1+\epsilon \end{aligned}$$

provided $I(A)$ is large enough. It then follows from (5.5), (5.9), (5.11), Lemma 3.1.2 and Theorem 3.1.10 that

$$\int_A e^{-I(x)}\mathrm{d}x \leqslant e^{-I(A)} \int_{\Lambda_{I(A)}} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|}\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\big(1+o(1)\big)$$

as $I(A)$ tends to infinity. Under (5.2), the set $\{\,\tau_A < \infty\,\}$ is included in ω_A. Consequently, (3.2.3) yields, as A tends to infinity,

$$\begin{aligned} \int_A e^{-I(x)}\mathrm{d}x \leqslant e^{-I(A)} \int_{u\geqslant 0} e^{-u} \int_{p\in\mathcal{D}_A} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|} \int_{v\in S_{T_p\pi_A^{-1}(p)}(0,1)} \\ \int_{t\in[0,e_A(p,v))} \frac{\mathrm{I}_{[0,u]}\Big(\tau_A\big(\exp_p(tv)\big)-\tau_A(p)\Big)}{\big|J\pi_A\big(\exp_p(tv)\big)\big|} \frac{|\mathrm{D}I(p)|}{\big|\mathrm{D}I\big(\exp_p(tv)\big)\big|}\times \\ \det\mathcal{A}(t,v)\,\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}u\,\big(1+o(1)\big)\,. \end{aligned}$$

Combine (5.8) and Proposition 4.2.4.i to upper $\det\mathcal{A}(t,v)$ from above. Use (5.9) to get an upper bound $|\mathrm{D}I(p)|/\big|\mathrm{D}I\big(\exp_p(tv)\big)\big|$ by $1+\epsilon$, and

(5.12) bound $\bigl|J\pi_A\bigl(\exp_p(tv)\bigr)\bigr|$ by $1+\epsilon$. Then, assumption (5.3) yields, for $I(A)$ large enough,

$$\int_A e^{-I(x)}\mathrm{d}x \leqslant e^{-I(A)}\int_{u\geqslant 0}\int_{p\in\mathcal{D}_A}\frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|}\int_{v\in S_{T_p\pi^{-1}(p)}(0,1)}\int_{t\in[0,e_A(p,v))} \mathrm{I}_{[0,u(1+\epsilon)]}\Bigl(\frac{1}{2}|\mathrm{D}I(p)|\langle G(p)tv,tv\rangle\Bigr)t^{d-k-2}\,\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}u\,(1+\epsilon)$$

Extend the integration in t over the domain $[0,+\infty)$ and use Proposition 4.3.1 to conclude

$$\int_A e^{-I(x)}\mathrm{d}x \leqslant e^{-I(A)}2^{\frac{d-k-1}{2}}\Gamma\Bigl(\frac{d-k+1}{2}\Bigr)\omega_{d-k-1}\times \int_{p\in\mathcal{D}_A}\frac{e^{-\tau_A(p)}}{|\mathrm{D}I|^{\frac{d-k+1}{2}}}\frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_A}}{(\det G_A)^{\frac{1}{2}}}(1+\epsilon)^{\frac{d-k+7}{2}}\,.$$

Since ϵ is arbitrary, this yields the proper upper bound since

$$2^{\frac{d-k-1}{2}}\Gamma\left(\frac{d-k-1}{2}\right)\omega_{d-k-1} = (2\pi)^{\frac{d-k-1}{2}}\,.$$

Before we can drop assumption (5.14), we need to prove the lower bound. To do so, apply Theorem 3.1.10, Lemma 3.1.6 with assumptions (5.5) and (5.10) to obtain, as A moves to infinity,

$$\int_A e^{-I(x)}\mathrm{d}x \geqslant e^{-I(A)}\int_{p\in\underline{A}_M}\frac{e^{-\tau_A(p)}}{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)}(p)\bigr)\bigr|}\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\bigl(1+o(1)\bigr)^{-2}. \tag{5.15}$$

Notice that

$$\begin{aligned}\frac{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)}(p)\bigr)\bigr|}{|\mathrm{D}I(p)|} &= \exp\Bigl(\frac{1}{2}\int_0^{\tau_A(p)}\frac{\mathrm{d}}{\mathrm{d}s}\log\bigl|\mathrm{D}I\bigl(\psi(p,s)\bigr)\bigr|^2\mathrm{d}s\Bigr)\\ &= \exp\Bigl(\frac{1}{2}\int_0^{\tau_A(p)}2\,\frac{\langle \mathrm{D}^2I\cdot N,N\rangle}{|\mathrm{D}I|^2}\bigl(\psi(p,s)\bigr)\mathrm{d}s\Bigr)\\ &\leqslant \exp\Bigl(\int_0^{\tau_A(p)}\frac{\|\mathrm{D}^2I\|}{|\mathrm{D}I|^2}\bigl(\psi_s(p)\bigr)\mathrm{d}s\Bigr)\,.\end{aligned}$$

This last inequality and assumption (5.11) imply

$$\lim_{A\to\infty}\sup\Bigl\{\Bigl|\frac{\bigl|\mathrm{D}I\bigl(\psi_{\tau_A(p)}(p)\bigr)\bigr|}{|\mathrm{D}I(p)|}-1\Bigr| : p\in\underline{A}_M\Bigr\} = 0\,.$$

Consequently, up to a multiplicative factor of $\big(1+o(1)\big)$, we can replace $\mathrm{D}I\big(\psi_{\tau_A(p)}(p)\big)$ by $\mathrm{D}I(p)$ in (5.15). We then use equality (3.2.3) and proceed as follows. First, we change the variable u into $\tau_A(p)+w$, and restrict the integration to w between 0 and w_0 for some positive w_0. Second, we restrict further the domain by integrating only over the points p in $\mathcal{D}_A$ with $\tau_A(p)$ less than $c_{A,M}-w_0$. On this range, we can use assumptions (5.9), (5.7), Proposition 4.2.4-ii, assumptions (5.12) and (5.3) to obtain, for $I(A)$ tending to infinity,

$$\int_A e^{-I(x)}\mathrm{d}x \geqslant e^{-I(A)} \int_{w\in[0,w_0]} e^{-w} \int_{\substack{p\in\mathcal{D}_A\\ \tau_A(p)\leqslant c_{A,M}-w_0}} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|}$$

$$\int_{v\in S_{T_p\pi^{-1}(p)}(0,1)} \int_{t\in[0,e_a(p,v))} \mathrm{I}_{[0,w(1-\epsilon)]}\Big(\frac{1}{2}|\mathrm{D}I(p)|\langle G_A(p)tv,tv\rangle\Big)$$

$$\mathrm{d}t\,\mathrm{d}\mu_p(v)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)\,\mathrm{d}w\big(1+o(1)\big)^{-2}.$$

Arguing as in the proof of Proposition 4.3.1, we obtain that for $I(A)$ large enough,

$$\int_A e^{-I(x)}\mathrm{d}x \geqslant e^{-I(A)}2^{\frac{d-k-1}{2}}\Gamma\Big(\frac{d-k+1}{2}\Big)\omega_{d-k-1}\times$$

$$\int_{\substack{p\in\mathcal{D}_A\\ \tau_A(p)\leqslant c_{A,M}-w_0}} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|^{\frac{d-k+1}{2}}\big(\det G_A(p)\big)^{\frac{1}{2}}}\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p)(1+\epsilon)^{-1}.$$

Since ϵ is arbitrary, assumption (5.13) gives then

$$\int_A e^{-I(x)}\mathrm{d}x$$
$$\geqslant e^{-I(A)}2^{\frac{d-k-1}{2}}\int_{p\in\mathcal{D}_A} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|^{\frac{d-k+1}{2}}\big(\det G_A(p)\big)^{\frac{1}{2}}}\,\mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p).$$

It remains for us to drop the assumption that $A=A\cap\Gamma_{I(A)+c_{A,M}}$ is the upper bound. This is immediate; for a general set A, write

$$\int_A e^{-I(x)}\mathrm{d}x = \int_{A\cap\Gamma_{I(A)+c_{A,M}}} e^{-I(x)}\mathrm{d}x + \int_{A\cap\Gamma^{\mathrm{c}}_{I(A)+c_{A,M}}} e^{-I(x)}\mathrm{d}x.$$

For the first integral in the right hand side of the above equality, the theorem — proved in this case! — gives the asymptotic equivalent. For the second one, it is less than the integral over $\Gamma^{\mathrm{c}}_{I(A)+c_{A,M}}$, that

is $L\big(I(A)+c_{A,M}\big)$. Assumption (5.4) shows that it has a negligible contribution to the asymptotics.

When $k=d-1$ and $\mathcal{D}_A$ is an open subset of $\Lambda_{I(A)}$, the asymptotic equivalent in Theorem 5.1 is nothing but (3.2.1). ∎

5.2. REMARK. Assumption (5.8) turns to be difficult to check in practice. The main reasons are that geodesics can seldom be explicitly calculated, and that the curvature tensor may be difficult to calculate. However, we only used it to apply Proposition 4.2.4.i when deriving the upper bound in the proof of Theorem 5.1. It would be enough to have

$$\begin{aligned}\lim_{A\to\infty}\sup\Big\{ \Big|\frac{\det\mathcal{A}(t,v)}{t^{d-k-2}}-1\Big| : v\in S_{T_p\pi^{-1}(p)}(0,1)\,,\\ t\in[\,0,e_A(p,v))\,,\ \exp_p(tv)\in\underline{A}_M\Big\}=0\,. \qquad (5.16)\end{aligned}$$

This condition will turn out to be easier to check in many cases. This could also replace (5.7) as well.

In a similar spirit, (5.3) may be tedious to verify. Often τ_A cannot be easily calculated, but is only known via an asymptotic expansion as A moves to infinity. Due to the error term in the asymptotic expansion, the uniformity in (5.3), for small q very close to p, may be difficult to check. Therefore, we will make a rather systematic use of the following weaker hypothesis. Assume that there exists a function $\tilde\tau_A$ defined on $\underline{A}_M$ such that

$$\lim_{A\to\infty}\sup_{q\in\underline{A}_M}|\tau_A(q)-\tilde\tau_A(q)|=0$$

and

$$\lim_{A\to\infty}\sup_{q\in\underline{A}_M}\Bigg|\frac{\tilde\tau_A(q)-\tilde\tau_A\big(\pi(q)\big)}{\frac12|\mathrm{D}I\big(\pi_A(q)\big)|\big\langle G_A\big(\pi_A(q)\big)\exp^{-1}_{\pi_A(q)}(q)\,,\exp^{-1}_{\pi_A(q)}\big\rangle}-1\Bigg|=0 \qquad (5.17)$$

for some $G(p)$ on $T_p\Lambda_{I(A)}$. Then, Theorem 5.1 holds when (5.3) is replaced by (5.17). Indeed, let ϵ be a positive number. In the proof of Theorem 5.1, we now use the bound

$$\begin{aligned}\mathrm{I}_{[0,u(1-\epsilon)-\epsilon]}&\Big(\frac12|\mathrm{D}I(p)|\langle G_A(p)tv,tv\rangle\Big)\\ &\leqslant \mathrm{I}_{[0,u]}\Big(\tau_A\big(\exp_p(tv)\big)-\tau_A(p)\Big)\\ &\leqslant \mathrm{I}_{[0,u(1+\epsilon)+\epsilon]}\Big(\frac12|\mathrm{D}I(p)|\langle G(p)tv,tv\rangle\Big)\,.\end{aligned}$$

By the dominated convergence theorem,

$$\lim_{\epsilon\to 0}\int_{u\geqslant 0} e^{-u}\bigl(u+(1\pm\epsilon)\pm\epsilon\bigr)^{\frac{d-k-1}{2}}\mathrm{d}u=\int e^{-u}u^{\frac{d-k-1}{2}}\mathrm{d}u\,,$$

and one easily sees that the proof of Theorem 5.1 goes through.

It is sometimes convenient to weaken even a tiny bit this assumption, only assuming that $\tau_A(q)-\tau_A\bigl(\pi_A(q)\bigr)$ can be approximated by some $\tilde\tau_A(q)-\tilde\tau_A\bigl(\pi_A(q)\bigr)$ in the sense that

$$\lim_{A\to\infty}\sup_{q\in\underline{A}_M}\Bigl|\tau_A(q)-\tau_A\bigl(\pi_A(q)\bigr)-\Bigl(\tilde\tau_A(q)-\tilde\tau_A\bigl(\pi_A(q)\bigr)\Bigr)\Bigr|=0$$

and of course keeping requirement (5.17).

Let us now explain how Theorem 5.1 can be used to obtain information on limiting conditional distributions.

Assume that we consider a log-concave density function proportional to e^{-I} on $\mathbb{R}^d$. As A moves away to infinity, there is not much hope for the conditional distribution

$$\int_{A\cap B} e^{-I(x)}\mathrm{d}x\Bigm/\int_A e^{-I(x)}\mathrm{d}x$$

to converge to a nontrivial limit. Indeed, a fixed bounded set B does not intersect A if $I(A)$ is large enough. So, we need to rescale B. For this, consider a normalizing function $A\mapsto\lambda_A\in(0,\infty)$. We will require that the dominating manifold $\mathcal{D}_A/\lambda_A$ converges in a weak sense. But for the time being, consider the rescaled conditional distribution

$$\mu_A(B)=\int_{A\cap\lambda_A B} e^{-I(x)}\mathrm{d}x\Bigm/\int_A e^{-I(x)}\mathrm{d}x\,.$$

It converges weakly* if for any continuous and bounded function f on $\mathbb{R}^d$, the integral

$$\int f\,\mathrm{d}\mu_A=\int_A f(x/\lambda_A)e^{-I(x)}\mathrm{d}x\Bigm/\int_A e^{-I(x)}\mathrm{d}x$$

converges as A moves to infinity. Its limit is a linear form in f, associated to a measure, the weak* limit of μ_A.

Let us assume that

as A moves to infinity, the family of rescaled measures

$$\frac{\lambda_A^k e^{-\tau_A(\lambda_A q)} \mathrm{d}\mathcal{M}_{\mathcal{D}_A/\lambda_A}(q)}{|\mathrm{D}I(\lambda_A q)|^{\frac{d-k+1}{2}} \big(\det G_A(\lambda_A q)\big)^{\frac{1}{2}}} \Big/ \Big/ \int_A e^{-\tau_A} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_A}}{|\mathrm{D}I|^{\frac{d-k+1}{2}} (\det G_A)^{\frac{1}{2}}}$$

converges weakly* to a probability measure ν. (5.18)

Assume furthermore that

$$\lim_{A\to\infty} \frac{c_{A,M}}{\lambda_A} \sup \{\, |\mathrm{D}I(p)|^{-1} \,:\, p \in \underline{A}_M \,\} = 0 \tag{5.19}$$

and

$$\lim_{A\to\infty} \sup \{\, |q-p|/\lambda_A \,:\, q \in \pi_A^{-1}(p)\,,\ \tau_A(p) \leqslant c_{A,M} \,\} = 0\,. \tag{5.20}$$

We then have the following convergence.

5.3. COROLLARY. *Under the assumptions of Theorem 5.1 and (5.15)–(5.20), the conditional distribution μ_A converges weakly* to ν as $I(A)$ tends to infinity.*

Proof. Argue as in the proof of Proposition 3.1.9 to obtain

$$\begin{aligned}
\int_A e^{-I(x)} f(x/\lambda_A)\,\mathrm{d}x &= \int\int \mathrm{I}_A(x)\mathrm{I}_{[I(x),\infty)}(c) e^{-c} f(x/\lambda_A)\,\mathrm{d}c\,\mathrm{d}x \\
&= e^{-I(A)} \int\int e^{-c}\, \mathrm{I}_{A\cap\Gamma_{I(A)+c}}(x) f(x/\lambda_A)\,\mathrm{d}c\,\mathrm{d}x \\
&= e^{-I(A)} \int_{c\geqslant 0} e^{-c} \int_{0\leqslant s\leqslant c} \int_{p\in\Lambda_{I(A)}} \mathrm{I}_A\big(\psi(p,s)\big) \\
&\qquad \Big|\frac{\mathrm{d}}{\mathrm{d}s}\psi(p,s)\Big| \det\big(\psi_{s*}^{\mathrm{T}}\psi_{s*}\big)^{1/2} f\big(\psi(p,s)/\lambda_A\big) \\
&\qquad \mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)\,\mathrm{d}s\,\mathrm{d}c\,.
\end{aligned}$$

Consider a function f, bounded and continuous. After adding a constant to f, we may assume that f is larger than some positive number. Then, up to introducing a term $f\big(\psi(p,s)/\lambda_A\big)$, the bounds in Theorem 3.1.10 remain valid.

Under (5.14) and arguing as in the proof of Theorem 5.1, as A tends to infinity, we have

$$\begin{aligned}
\int_A e^{-I(x)} f(x/\lambda_A)\,\mathrm{d}x \sim e^{-I(x)} &\int_{\substack{p\in\Lambda_{I(A)}\\ \tau_A(p)<\infty}} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|} \times \\
&\int_{0\leqslant s\leqslant\chi_A(p)} e^{-s} f\big(\psi_{\tau_A(p)+s}(p)/\lambda_A\big)\,\mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p)
\end{aligned}$$

where $\chi_A^F(p) \leqslant \chi_A(p) \leqslant \chi_A^L(p)$. Assumption (5.14), (5.19) and Corollary 3.1.4 imply

$$\frac{1}{\lambda_A}|\psi_{\tau_A(p)+s}(p) - p| \leqslant \frac{\chi_A(p)}{\lambda_A|\mathrm{D}I(p)|} \leqslant \frac{c_{A,M}}{\lambda_A|\mathrm{D}I(p)|} = o(1) \tag{5.21}$$

as $I(A)$ tends to infinity, uniformly in $p \in \Lambda_{I(A)}$ with $\tau_A(p) < \infty$.

Assume that f is uniformly continuous. Since f is bounded and larger than some positive number, (5.21) implies

$$\lim_{A\to\infty} \sup\Big\{ \Big|\frac{f\big(\psi_{\tau_A(p)+s}(p)/\lambda_A\big)}{f(p/\lambda_A)} - 1\Big| \, : \, p \in \underline{A}_M \, , \, 0 \leqslant s \leqslant c_{A,M} \Big\} = 0 \, .$$

Thus, using (5.5),

$$\int_A e^{-I(x)} f(x/\lambda_A) \mathrm{d}x \sim e^{-I(A)} \int_{\substack{p\in\Lambda_{I(A)} \\ \tau_A(p)<\infty}} \frac{e^{-\tau_A(p)}}{|\mathrm{D}I(p)|} f(p/\lambda_A) \, \mathrm{d}\mathcal{M}_{\Lambda_{I(A)}}(p) \, .$$

We make a change of variable as we did in (3.2.2). Noting that (5.20) implies

$$\lim_{A\to\infty} \sup\Big\{ \Big|\frac{f\big(\exp_p(v)/\lambda_A\big)}{f(p/\lambda_A)} - 1\Big| \, : \, v \in T_p\pi_A^{-1}(p) \, , \\ \tau_A\big(\exp_p(v)\big) \leqslant c_{A,M} \, , \, p \in \mathcal{D}_A \Big\} = 0 \, ,$$

we obtain

$$\int_A e^{-I(x)} f(x/\lambda_A) \mathrm{d}x \sim e^{-I(A)} (2\pi)^{\frac{d-k-1}{2}} \times \\ \int_{\mathcal{D}_A} \frac{e^{\tau_A(p)} f(p/\lambda_A)}{|\mathrm{D}I(p)|^{\frac{d-k+1}{2}} \big(\det G_A(p)\big)^{\frac{1}{2}}} \, \mathrm{d}\mathcal{M}_{\mathcal{D}_A}(p) \, .$$

Make a change of variable $q = p/\lambda_A$ and use (5.18) to obtain

$$\lim_{A\to\infty} \frac{\int_A e^{-I(x)} f(x/\lambda_A) \, \mathrm{d}x}{\int_A e^{-I(x)} \mathrm{d}x} = \int f \mathrm{d}\nu \, , \tag{5.22}$$

for all uniformly continuous, positive functions f.

If f is bounded, we drop the restriction (5.14), as we did in the proof of Theorem 5.1. Using Theorem 5.1, we can consider arbitrary bounded uniformly continuous function f in (5.22). This implies —

see, e.g., Pollard (1984) — that the conditional distribution converges weakly*. ∎

Notes

There are many things related to Theorem 5.1 that I wanted to do but could not.

A first one is to understand to what extent an exponentially integrable density may be approximated by a log-concave one at infinity. Here is the beginning of what could be a proof. Let f be a density on $\mathbb{R}^d$, such that the moment generating function

$$\phi(t) = \int e^{\langle t,x\rangle} f(x)\mathrm{d}x$$

is finite in a neighborhood of the origin. Under some classical steepness conditions — see, e.g., Barndorff-Nielsen (1978) or Brown (1986) — the differential $m = \mathrm{D}\phi$ is a diffeomorphism. Denote by m^{inv} its inverse. We now follow word for word the construction of Barbe and Broniatowski (200?), but in a different setting.

The function $\log\phi$ is convex. Let I be its convex conjugate, that is

$$I(x) = \sup\{\,\log\phi(t) - \langle t,x\rangle\,\}\,.$$

Using the change of variable $a = s-x$ and Fubini's theorem, we obtain

$$\int\int e^{\langle m^{\mathrm{inv}}(s-x),x\rangle} \mathrm{I}_A(s-x) f(s)\,\mathrm{d}s\,\mathrm{d}x = \int_A e^{-I(a)}\mathrm{d}a\,.$$

This allows us to define a new density

$$g_A(x) = \frac{\int e^{\langle m^{\mathrm{inv}}(s-x),x\rangle} \mathrm{I}_A(s-x) f(s)\mathrm{d}s}{\int_A e^{-I(a)}\mathrm{d}a}\,.$$

The interesting fact is that $g_A(0) = \int_A f(s)\mathrm{d}s \Big/ \int_A e^{-I(a)}\mathrm{d}a$. Consider the rescaled density $r^d g_A(rx)$ with r possibly depending on A. If we can prove that $r^d g_A(r\,\cdot)$ converges to a limit, say $g(\cdot)$, as A moves to infinity, in such a way that pointwise convergence at 0 holds, then

$$\int_A f(s)ds \sim r^{-d} g(0) \int_A e^{-I(a)}\mathrm{d}a\,.$$

Thus, when integrating over A, we can approximate the density f by a multiple of $e^{-I(A)}$, and then use Theorem 5.1.

To achieve this approximation, we can calculate the Fourier transform of $r^d g_A(r\,\cdot)$. It is

$$\hat{g}_{r,A}(\lambda) = r^d \int e^{-\langle\lambda,x\rangle} g_A(rx)\,\mathrm{d}x = \int e^{-k(a,\lambda/r)}\,\mathrm{d}\mu_A(a)\,,$$

with

$$k(a,\lambda) = i\langle a,\lambda\rangle + \log\phi\bigl(m^{\mathrm{inv}}(a)\bigr) - \log\phi\bigl(m^{\mathrm{inv}}(a)+i\lambda\bigr)\,,$$

and

$$\mu_A(B) = \int_{A\cap B} e^{-I(a)}\mathrm{d}a \Big/ \int_A e^{-I(a)}\mathrm{d}a\,.$$

In particular, as A moves to infinity, the support of μ_A — that is A — moves to infinity. A Taylor expansion of $k(a,\lambda/r)$ near $\lambda/r=0$ gives

$$k(a,\lambda/r) = \frac{1}{2}\Bigl\langle \frac{(\mathrm{D}^2\log\phi)\circ m^{\mathrm{inv}}(a)}{r^2}\lambda,\lambda\Bigr\rangle + o(\lambda^2)\,.$$

If this can be done as a tends to infinity — which is in the spirit of what we did using Proposition 4.1.1 to prove Theorem 5.1 — we can hope to approximate

$$\hat{g}_{r,A}(\lambda) \approx \int \exp\Bigl(\frac{1}{2}\Bigl\langle \frac{(\mathrm{D}^2\log\phi)\circ m^{\mathrm{inv}}(a)}{r^2}\lambda,\lambda\Bigr\rangle\Bigr)\mathrm{d}\mu_A(a)\,.$$

Let $V(a) = (\mathrm{D}^2\log\phi)\circ m^{\mathrm{inv}}(a)$ be the so-called variance function of f. Inverting the Fourier transform of the approximation, we should obtain

$$r^d g(0) \approx \int \frac{1}{(2\pi)^{d/2}\det(V(a)/r)^{1/2}}\mathrm{d}\mu_A(a)\,.$$

If we can find r depending on A such that the right hand side has a limit as A tends to infinity, and use Corollary 5.3, we are done.

Unfortunately, I could not come up with useful conditions for this idea to work.

When $d=1$, it is possible to prove that if g_A converges, then its limit is given by a mixture of either normal densities — as we outlined here — or gamma ones; this follows from Balkema, Klüppelberg and Resnick (1999). In higher dimensions, a related approximation is in Barndorff-Nielsen and Klüppelberg (1999). I somewhat believe that the whole virtue of saddlepoint approximations used in statistics is to provide some form of log-concave approximation in the spirit of what is outlined here. But most of the time, in the multivariate setting, it

relies on assumptions similar to the convergence of g_A, which I don't find too appealing. I have been searching unsuccessfully for a decent condition on f itself.

A second project, which perhaps would be desirable to carry out, is to obtain higher order expansions. Theorem 5.1 provides a one term asymptotic expansion. Starting from the equality in Proposition 3.1.9, one could do the change of variable using the dominating manifold and the orthogonal leaves; then one would use asymptotic expansions for whatever function is involved, and obtain the desired approximation. Higher order differential geometry is involved. The difficulty is to come up with a set of usable conditions to perform all the approximations. Another route would be to mimic the practice of Edgeworth expansions in statistics. There are essentially two types of them: those that are proved rigorously, and a vast majority that are called "formal". To do the formal ones, the argument is pretty much to neglect what one believes to be negligible under some quite unknown conditions, and proceed. One could obtain formal asymptotic expansion in the same way. This may be of some value in a few applications. Indeed, sometimes one may not look for a theorem but maybe more for a guideline.

A third path to explore would be to derive explicit upper bounds, starting either from Proposition 3.1.9 or Theorem 3.1.10. In particular, I wonder if the technique developed here could be of any use to investigate "asymptotic" isoperimetric problems.

As pointed out in the notes to chapter 3, Proposition 3.1.9 does not use the convexity of I. This proposition is true for any smooth function I for which the sets Λ_c are smooth hypersurfaces. There may be some examples where the normal flow can be calculated explicitly and other arguments used in order to derive an estimate similar to that of Theorem 5.1.

To conclude these notes, Corollary 5.3 is inspired by the Gibbs conditioning principle in large deviations. In the large deviation context, the reader may consult Csiszár (1984) and Bolthausen (1993)

6. Asymptotics for sets translated towards infinity

In this section, we study integrals of the form $\int_{A+t} e^{-I(x)}\mathrm{d}x$ as $|t|$ tends to infinity. To avoid any ambiguity, recall that if A is a set and t is a vector, both in $\mathbb{R}^d$, the translation of A by t is

$$A+t = \{\, x+t : x \in A \,\}\,.$$

We will assume that

$$\text{$A \subset \mathbb{R}^d$ is a closed bounded convex neighborhood of the origin, with smooth boundary and positive curvature.} \tag{6.1}$$

The only restriction here is convexity — and smoothness, but we want to be able to use differential geometric methods! It could be dropped at the cost of a more sophisticated discussion on how $A+t$ and $\Lambda_{I(A+t)}$ intersect. Up to changing t by a fixed amount, we can always assume that A contains the origin.

We will control the growth of I at infinity, assuming that

$$\lim_{p\to\infty} \frac{\log I(p)}{|\mathrm{D}I(p)|} = 0 \qquad \text{and} \qquad \lim_{p\to\infty} \frac{\|\mathrm{D}^2 I(p)\|}{|\mathrm{D}I(p)|} = 0\,. \tag{6.2}$$

The second condition forces I not to increase too fast. Indeed, for $d=1$, it reads $|I''(p)/I'(p)|$ tends to 0 as $|p|$ tends to infinity. Hence, for any small positive ϵ and any p, q large enough,

$$\Big|\frac{I'(p)}{I'(q)}\Big| = \Big|\exp\int_q^p \frac{I''(t)}{I'(t)}\mathrm{d}t\Big| \leqslant \exp\big(\epsilon|p-q|\big)\,.$$

For instance, the function $I(p)=\exp(|p|^\alpha)$ satisfies $(I''/I')(p)$ tends to 0 as p tends to infinity if and only if $\alpha<1$. So, roughly, I should have a subexponential growth. A polynomial growth, like $I(p)=|p|^\alpha$ with a positive α, is admissible.

The first condition forces I to increase fast enough. Indeed, when $d=1$, it implies that for any positive ϵ and $x<y$ large enough,

$$I(y)-I(x) = \int_x^y I'(t)\mathrm{d}t \geqslant \frac{1}{\epsilon}\int_x^y \log I(t)\mathrm{d}t \geqslant (y-x)\frac{\log I(y)}{\epsilon}\,.$$

Hence, ultimately, I has to grow faster that any linear function. For instance, the function $I(t) = t(\log t)^\alpha$ satisfies $\big(\log I(t)\big)/I'(t)$ tends to 0 as t tends to infinity if and only if $\alpha > 1$.

We will need to strengthen the second condition in (6.2) by assuming

$$\lim_{t\to\infty} \log I(A+t) \sup\Big\{ \frac{\|\mathrm{D}^2 I(q)\|}{|\mathrm{D}I(q)|} : I(q) \geqslant I(A+t) \Big\} = 0\,. \tag{6.3}$$

For a strictly convex function I the level sets Γ_c are strictly convex. There is a unique point x in $\mathbb{R}^d$ at which I is minimal. Assumption (6.1) implies that for any t with $|t| > 2\,\mathrm{diam}(A) + |x|$,

$$(A+t) \cap \Lambda_{I(A+t)} = \{\, p_t \}$$

for a unique point p_t. In particular, $p_t - t$ is in ∂A.

6.1. THEOREM. *If I is convex and (6.1)–(6.3) hold, then*

$$\int_{A+t} e^{-I(x)} \mathrm{d}x \sim \frac{(2\pi)^{\frac{d-1}{2}}}{|\mathrm{D}I(p_t)|^{\frac{d+1}{2}} K_t}\, e^{-I(A+t)} \qquad \text{as } t \to \infty\,,$$

where K_t is the Gauss-Kronecker curvature of ∂A at $p_t - t$.

Proof. Write $A_t = A + t$. In order to apply Theorem 5.1, we need to have a candidate for the dominating manifold $\mathcal{D}_{A_t}$. Clearly $\{\, p_t\}$ should do, and we set $\mathcal{D}_{A_t} = \{\, p_t\}$. Since $\tau_{A+t}(p_t) = 0$ by definition of p_t, the result of Theorem 5.1 reads

$$\int_{A+t} e^{-I(x)} \mathrm{d}x \sim \frac{e^{-I(A+t)} (2\pi)^{\frac{d-1}{2}}}{|\mathrm{D}I(p_t)|^{\frac{d+1}{2}} \big(\det G_{A+t}(p_t)\big)^{\frac{1}{2}}} \qquad \text{as } t \to \infty\,.$$

Now, recall that we should expect $G_{A+t}(p_t)$ to be the difference of the second fundamental forms of $\Lambda_{I(A+t)}$ and $A+t$ at p_t — not restricted to anything here, since $\mathcal{D}_{A_t}$ is a point and so $\pi_{A_t}^{-1}(p_t)$ is $\Lambda_{I(A+t)}$, up to what is in the cut locus of p_t. However, the second part of assumption (6.2) asserts that asymptotically, the second fundamental form of $\Lambda_{I(A+t)}$ degenerates, and so, locally, $\Lambda_{I(A+t)}$ is almost flat. Thus, $G_{A+t}(p_t)$ should be the second fundamental form of $\partial(A+t)$ at p_t, which is equal to that of ∂A at $p_t - t$. Its determinant is exactly K_t. This explains how to guess the result. It is hoped that this twelve line argument convinces the reader that Theorem 5.1 can be useful.

Now that the result is guessed, let us find a candidate for $c_{A+t,M}$. Define

$$c(t) = d \log I(A+t) + \frac{d+2}{2} \log |\mathrm{D}I(p_t)|$$

— in this formula, d refers to the dimension of $\mathbb{R}^d$. Since $I(p_t)$ tends to infinity with t, the first part of (6.2) implies $\lim_{t\to\infty} |\mathrm{D}I(p_t)| = \infty$. Proposition 2.1 yields

$$L\big(I(A+t) + c(t)\big) = \frac{e^{-I(p_t)}}{|\mathrm{D}I(p_t)|^{\frac{d+1}{2}}} o(1) \qquad \text{as } t \to \infty\,.$$

Given (5.4) and our twelve line argument, $c_{A+t,M} = c(t)$ is a good candidate, no matter what M is. It guarantees (5.4) as well as (5.6).

We now check all the assumptions of Theorem 5.1.

As noted in chapter 5, since $\mathcal{D}_{A_t} = \{\, p_t \}$ is a point, (5.1) is trivial. Notice that $\underline{A}_{t,M}$ is included in the projection of $A+t$ on $\Lambda_{I(A+t)}$. So it is enough to check (5.2) with $\underline{A}_{t,M}$ replaced by the projection of $A+t$. Assumption (6.2) asserts that the second fundamental form of $\Lambda_{I(A+t)}$ tends to 0 uniformly over this surface. Thus, its curvature tensor vanishes asymptotically and the radius of injectivity of any point in $\Lambda_{I(A+t)}$ tends to infinity uniformly over the surface — this follows from Rauch's (1951) theorem or Klingenberg's (1959) lemma; see, e.g., Do Carmo (1992) or Chavel (1996). As $A+t$ stays of finite diameter, (5.2) follows.

To prove (5.3) and find G_{A+t}, we need to have some more information on $\underline{A}_{t,M}$ and on the normal flow. The idea is that $\underline{A}_{t,M}$ should be very close to $\mathcal{D}_{A_t} = \{\, p_t \}$. To prove this fact, we first define a family of local parameterizations of ∂A. For p belonging to ∂A, we denote by $\nu(\cdot)$ the inward unit normal vector to ∂A at p. By compactness of A, there exists a positive ϵ_1, independent of p, such that $\partial A \cap \mathcal{B}(p, \epsilon_1)$ can be parametrized as all points of the form $p + u + f_p(u)\nu(p)$ for u in $T_p\partial A$ and some nonnegative function f_p. Notice that the curvature assumption (6.1) ensures that there exists a positive matrix Q_p such that $f_p(u) = \frac{1}{2}\langle Q_p u, u\rangle\big(1 + o(1)\big)$ as $|u|$ tends to 0. Moreover, since ∂A is smooth and compact, the term $o(1)$ is uniform when p varies in ∂A, and the matrices Q_p are bounded bellow by a fixed positive one.

To prove that $\underline{A}_{t,M}$ shrinks around p_t, notice that $\nu(p_t - t)$ and $\mathrm{D}I(p_t)$ are collinear since $\partial(A+t)$ and $\Lambda_{I(A+t)}$ are tangent at p_t.

Convexity of I implies

$$\begin{aligned}
&I\big(p_t+u+f_{p_t-t}(u)\nu(p_t-t)\big)-I(p_t)\\
&\qquad\geqslant \frac{\mathrm{d}}{\mathrm{d}s}I\Big(p_t+s\big(u+f_{p_t-t}(u)\nu(p_t-t)\big)\Big)\Big|_{s=0}\\
&\qquad= f_{p_t-t}(u)|\mathrm{D}I(p_t)|\,.
\end{aligned}$$

Consequently, the points q in $\partial A+t$ such that $I(q)\leqslant I(A+t)+c_{A+t,M}$ can be parametrized as $p_t+u+f_{p_t}(u)\nu(p_t-t)$ with $|u|^2\leqslant O\big(c(t)/|\mathrm{D}I(p_t)|\big)=o(1)$. They can also be written as

$$p_t+u+\frac{1}{2}\langle Q_{p_t-t}u,u\rangle\nu(p_t-t)\big(1+o(1)\big) \tag{6.4}$$

where the $o(1)$ is uniform in t and $|u|^2\leqslant O\big(c(t)/|\mathrm{D}I(p_t)|\big)$.

Since the curvature of the level set of I tends to 0, the normal flow should be almost like straight lines on sizeable intervals. In order to make this statement rigorous, and seeking a linear approximation of the normal flow with good error bounds, an elementary calculation shows that

$$\mathrm{D}\Big(\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\Big)=\frac{1}{|\mathrm{D}I|}\Big(\frac{\mathrm{D}^2I}{|\mathrm{D}I|}-2N\otimes N\frac{\mathrm{D}^2I}{|\mathrm{D}I|}\Big)=\frac{1}{|\mathrm{D}I|}(\mathrm{Id}-2N\otimes N)\frac{\mathrm{D}^2I}{|\mathrm{D}I|}\,.$$

In particular, this implies the inequality

$$\Big\|\mathrm{D}\Big(\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\Big)\Big\|\leqslant\frac{1}{|\mathrm{D}I|}\frac{\|\mathrm{D}^2I\|}{|\mathrm{D}I|}\,. \tag{6.5}$$

Notice also that assumption (6.2) insures that

$$\eta(t)=\sup\Big\{\frac{\|\mathrm{D}^2I(q)\|}{|\mathrm{D}I(q)|}:I(q)\geqslant I(p_t)\Big\}=o(1)\qquad\text{as } t\to\infty\,.$$

Using Lemma 3.1.1 and (6.5), it follows that with $q=\exp_{p_t}(u)$,

$$\begin{aligned}
\Big|\psi(q,s)-q-s\frac{\mathrm{D}I}{|\mathrm{D}I|^2}(q)\Big|&=\Big|\int_0^s\int_0^r\frac{\mathrm{d}}{\mathrm{d}v}\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\big(\psi(q,v)\big)\,\mathrm{d}v\,\mathrm{d}r\Big|\\
&\leqslant\frac{\eta(t)}{|\mathrm{D}I(q)|}\frac{s^2}{2}\,.
\end{aligned} \tag{6.6}$$

Next, let us prove that $|\mathrm{D}I(q)|\sim|\mathrm{D}I(p_t)|$ as t tends to infinity provided u stays bounded. Writing $\gamma(s)=\exp_{p_t}\big(su/|u|\big)$,

$$\begin{aligned}
\log|\mathrm{D}I(q)|-\log|\mathrm{D}I(p_t)|&=\int_0^{|u|}\frac{\mathrm{d}}{\mathrm{d}s}\log\big|\mathrm{D}I\big(\gamma(s)\big)\big|\,\mathrm{d}s\\
&=\int_0^{|u|}\frac{1}{2}\frac{\big\langle\mathrm{D}^2I\big(\gamma(s)\big)\gamma'(s)\,,\mathrm{D}I\big(\gamma(s)\big)\big\rangle}{\big|\mathrm{D}I\big(\gamma(s)\big)\big|^2}\,\mathrm{d}s\,,
\end{aligned}$$

from which we obtain the bound

$$\bigl|\log|\mathrm{D}I(q)| - \log|\mathrm{D}I(p_t)|\bigr| \leqslant \frac{|u|}{2}\eta(t) = \frac{|u|}{2}o(1) \qquad \text{as } t \to \infty \tag{6.7}$$

Furthermore, we have a good control on the oscillations of $\mathrm{D}I/|\mathrm{D}I|^2$ in using (6.5); namely, for t large enough and say $|u| \leqslant 1$,

$$\begin{aligned}\Bigl|\frac{\mathrm{D}I}{|\mathrm{D}I|^2}(q) - \frac{\mathrm{D}I}{|\mathrm{D}I|^2}(p_t)\Bigr| &= \int_0^{|u|} \frac{\mathrm{d}}{\mathrm{d}s}\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\bigl(\gamma(s)\bigr)\mathrm{d}s \\ &\leqslant \frac{2|u|}{|\mathrm{D}I(p_t)|}\eta(t)\,.\end{aligned}$$

Consequently, for $|u| \leqslant 1$, the inequality (6.6) gives the bound

$$\Bigl|\psi(q,s) - q - s\frac{\mathrm{D}I}{|\mathrm{D}I|^2}(p_t)\Bigr| \leqslant \frac{2\eta(t)}{|\mathrm{D}I(p_t)|}\bigl(s^2 + s|u|\bigr)\,.$$

This is the linear approximation of the normal flow that we were looking for. Considering $s = \tau_{A+t}(q)$ and using the linear approximation for the exponential map in Proposition A.2.1 — remember that $q = \exp_{p_t}(u)$ — we then obtain

$$\begin{aligned}\psi\bigl(q,\tau_{A+t}(q)\bigr) &= p_t + u + \tau_{A+t}(q)\frac{N(p_t)}{|\mathrm{D}I(p_t)|} \\ &\quad + \frac{\eta(t)}{|\mathrm{D}I(p_t)|}\bigl(\tau_{A+t}(q)^2 + \tau_{A+t}(q)|u|\bigr)O(1) + \eta(t)|u|^2O(1)\end{aligned}$$

where the $O(1)$-terms are uniform in $|u| \leqslant 1$ as $|t|$ tends to infinity. Since $\psi\bigl(q,\tau_{A+t}(q)\bigr)$ is in the boundary of $A+t$ by the very definition of $\tau_{A+t}(q)$, and since u belongs to $T_{p_t}\Lambda_{I(p_t)}$, (6.4) forces us to have

$$\begin{aligned}\tau_{A+t}(q)\frac{N(p_t)}{|\mathrm{D}I(p_t)|} + \frac{\eta(t)}{|\mathrm{D}I(p_t)|}\bigl(\tau_{A+t}(q)^2 + \tau_{A+t}(q)|u|\bigr)O(1) + \eta(t)|u|^2O(1) \\ = \frac{1}{2}\langle Q_{p_t-t}u,u\rangle\nu(p_t - t)\bigl(1+o(1)\bigr)\end{aligned}$$

as t tends to infinity, and uniformly in $|u| \leqslant 1$. Therefore, provided $\eta(t)\tau_{A+t}(q) = o(1)$, we obtain

$$\tau_{A+t}(q) = \frac{|\mathrm{D}I(p_t)|}{2}\langle Q_{p_t-t}u,u\rangle\bigl(1+o(1)\bigr)$$

uniformly in $|u| \leqslant 1$, as t tends to infinity. Since $\eta(t)\tau_{A+t}(q) \leqslant \eta(t)c(t)$, assumption (6.3) ensures that $\eta(t)\tau_{A+t}(q) = o(1)$, and we

proved that (5.3) holds with G_{A+t} being the second fundamental form of ∂A at $p_t - t$.

Given our checking of (5.3), (5.5) is obvious since $\underline{A}_{t,M}$ shrinks around p_t — see the proof that $|u|^2 = o(1)$ before equation (6.4) — and (5.5) holds.

Assumption (5.7) is trivially satisfied. The shrinking of $\underline{A}_{t,M}$ to $\{\,p_t\}$ and assumption (6.2) — which implies that the curvature tends to 0; see also Proposition 4.2.3 — imply that $t_{0,M}(p) = o(1)$ and $K_{\max}\big(p, t_{0,M}(p)\big) = o(1)$ uniformly over $\underline{A}_{t,M}$ as t tends to infinity.

Assumption (5.8) is satisfied since I is convex, $\mathcal{D}_{A+t}$ is a point, and the first part of Proposition 4.2.3 holds.

Assumption (5.9) follows from (6.7) and the shrinking of $\underline{A}_{t,M}$ around p_t.

Clearly, (6.2) implies (5.10).

Assumption (5.11) is implied by (6.3) and Lemma 3.1.2, while (5.12) holds systematically for $k = 0$.

Since τ_{A+t} vanishes on $\mathcal{D}_{A+t}$, (5.13) holds as well, and this concludes the proof of Theorem 6.1. ∎

We obtained the conclusion of Theorem 6.1 by a brute application of Theorem 5.1. A little extra work makes the asymptotic formula nicer, replacing the term $|\mathrm{D}I(p_t)|$ by $|\mathrm{D}I(t)|$.

6.2. COROLLARY. *Under the assumptions of Theorem 6.1,*

$$\int_{A+t} e^{-I(x)}\mathrm{d}x \sim \frac{(2\pi)^{(d-1)/2}}{|\mathrm{D}I(t)|^{(d+1)/2} K_t} e^{-I(A+t)} \qquad \textit{as } t \to \infty\,.$$

Proof. Since

$$\frac{\mathrm{d}}{\mathrm{d}s} \log |\mathrm{D}I(t+su)|^2 = 2\frac{\langle \mathrm{D}^2 I(t+su)u\,, \mathrm{D}I(t+su)\rangle}{|\mathrm{D}I(t+su)|^2}\,,$$

the inequality

$$\big|\log |\mathrm{D}I(t+u)|^2 - \log |\mathrm{D}I(t)|^2\big| \leqslant 2\int_0^1 \frac{\|\mathrm{D}^2 I(t+su)\|}{|\mathrm{D}I(t+su)|} \frac{|u|}{|\mathrm{D}I(t+su)|}\,\mathrm{d}s$$

holds. Compactness of A, convexity of I and (6.2) imply that the right hand side of the above inequality is $o(1)$, uniformly in u belonging to A as t tends to infinity. Thus, $|\mathrm{D}I(t+u)| \sim |\mathrm{D}I(t)|$ uniformly over u

in A as t tends to infinity; and we can replace $|\mathrm{D}I(p_t)|$ by $|\mathrm{D}I(t)|$ in the statement of Theorem 6.1. ∎

In general, we do not have $I(A+t)-I(t)=o(1)$ as t tends to infinity. This is easily seen when $d=1$ and $I(x)=x^2$ for instance. Thus we cannot replace $I(A+t)$ by $I(t)$ in Theorem 6.1 or Corollary 6.2. In some instances, it is possible to obtain an asymptotic expansion for $I(A+t)$. We illustrate this fact in the important case when I is α-positively homogeneous. To state the result, recall the notation $N=\mathrm{D}I/|\mathrm{D}I|$ for the outward unit normal vector field to the level lines of I, and set $\Pi=\mathrm{D}^2I/|\mathrm{D}I|$. The compression of Π to the tangent space of a level line Λ_c is its second fundamental form.

6.3. PROPOSITION. *Assume that I is α-positively homogeneous and smooth. Assume also that A is a neighborhood of the origin with a smooth boundary. Let e be a unit vector in $\mathbb{R}^d$. Define r by the condition $-rN(e)\in\partial A$. Then, as λ tends to infinity, $I(\lambda e+A)$ admits an asymptotic expansion over the powers $\lambda^{\alpha-i}$, $i\in\mathbb{N}$, and*

$$\begin{aligned}I(A+\lambda e)=\lambda^\alpha I(e)-\lambda^{\alpha-1}r|\mathrm{D}I(e)|+\frac{\lambda^{\alpha-2}}{2}r^2\langle\Pi N(e)\,,N(e)\rangle\\ -\frac{\lambda^{\alpha-3}}{6}r^3\mathrm{D}^3I(e)\big(N(e)\,,N(e)\,,N(e)\big)\\ -\frac{\lambda^{\alpha-3}}{2}r^2\langle\Pi(e)Q^{-1}_{-rN(e)}\Pi(e)^{\mathrm{T}}N(e)\,,N(e)\rangle+O(\lambda^{\alpha-4})\,.\end{aligned}$$

Proof. Since A is compact and I is smooth and α-positively homogeneous, we have, uniformly in u belonging to A and as λ tends to infinity,

$$\begin{aligned}I(\lambda e+u)&=\lambda^\alpha I(e+u/\lambda)\\&=\lambda^\alpha I(e)+\sum_{1\leqslant i\leqslant k}\frac{\lambda^{\alpha-i}}{i!}\mathrm{D}^iI(e)\underbrace{(u\,,\ldots,u)}_{i\text{ times}}+O(\lambda^{\alpha-k-1})\,.\end{aligned}$$

The expansion of $I(\lambda e+A)$ follows by induction. The computation of the first terms can be done by introducing the point $u_\lambda=u(\lambda,e)$ in A, such that $I(e+A/\lambda)=I(e+u_\lambda/\lambda)$. Since A is convex, compact, and I is convex, u_λ in ∂A for λ large enough.

Taylor's expansion gives, uniformly in u belonging to A,

$$\begin{aligned}I\Big(e+\frac{u}{\lambda}\Big)=I(e)+\frac{|\mathrm{D}I(e)|}{\lambda}\langle N(e)\,,u\rangle+\frac{|\mathrm{D}I(e)|}{2\lambda^2}\langle\Pi(e)u\,,u\rangle\\+\frac{1}{6\lambda^3}\mathrm{D}^3I(e)(u\,,u\,,u)+O(\lambda^{-4})\end{aligned}$$

as λ tends to infinity. Consequently, $u_\lambda = -rN(e) + u_{1,\lambda}/\lambda$ where $u_{1,\lambda} = O(1)$ as λ tends to infinity — because u_λ has to minimize $I(e+u/\lambda)$, and so should minize $\langle N(e), u\rangle$ as well, up to a term of order λ^{-2}. Since $e + u_\lambda/\lambda$ belongs to ∂A, using the notation of the proof of Theorem 6.1, there exists a vector $v = v_\lambda$ in $T_{-rN(e)}\partial A = N(e)^\perp$ such that

$$
\begin{aligned}
\frac{u_{1,\lambda}}{\lambda} &= \frac{v}{\lambda} + f_{-rN(e)}\Big(\frac{v}{\lambda}\Big)N(e)\\
&= \frac{v}{\lambda} + \frac{1}{2\lambda^2}\langle Q_{-rN(e)}v, v\rangle N(e) + O(\lambda^{-3})
\end{aligned}
$$

as λ tends to infinity. It follows that

$$
\begin{aligned}
I\Big(e + \frac{u_\lambda}{\lambda}\Big) = I(e) - \frac{r|\mathrm{D}I(e)|}{\lambda} &+ \frac{r^2}{2\lambda^2}|\mathrm{D}I(e)|\langle \Pi(e)N(e)\,, N(e)\rangle\\
&+ \frac{|\mathrm{D}I(e)|}{2\lambda^3}\langle Q_{-rN(e)}v\,, v\rangle - \frac{r}{\lambda^3}|\mathrm{D}I(e)|\langle \Pi(e)v\,, N(e)\rangle\\
&- \frac{r^3}{6\lambda^3}\mathrm{D}^3 I(e)\big(N(e)\,, N(e)\,, N(e)\big) + O(\lambda^{-4})\,.
\end{aligned}
$$

The term in $1/\lambda^3$ is smallest when $v = rQ^{-1}_{-rN(e)}\Pi(e)^T N(e)$ — we used that Q is symmetric — and its minimum value is

$$
\begin{aligned}
-\frac{|\mathrm{D}I(e)|}{2\lambda^3}r^2\langle \Pi(e)Q^{-1}_{-rN(e)}\Pi(e)^T N(e)\,, N(e)\rangle&\\
- \frac{r^3}{6\lambda^2}\mathrm{D}^3 I(e)\big(N(e)\,, N(e)\,, N(e)\big)&\,.
\end{aligned}
$$

This completes the proof. ■

In particular, for $\alpha = 2$, we can replace $I(\lambda e + A)$ in the exponential term of the asymptotic equivalent by

$$
\lambda^2 I(e) - \lambda r|\mathrm{D}I(e)| + \frac{r^2}{2}\langle \Pi(e)N(e)\,, N(e)\rangle\,.
$$

In the Gaussian setting, $I(x) = -|x|^2/2$. Thus, $|DI(e)| = 1$ and $\Pi(c) = \mathrm{Id}$. The exponential term simplifies to

$$
-\frac{\lambda^2}{2} - \lambda r + \frac{r^2}{2}\,.
$$

A neat expression, but very specific to the Gaussian distribution…

Similarly to what we did in Corollary 5.2, we can obtain a result on conditional distribution. It is easy to prove that if X is a random variable with density e^{-I}, then, the conditional distribution of $X/|t|$ given $X \in A+t$ can be approximated by a point mass at $t/|t|$, under the assumptions of Theorem 6.1. However, Theorem 6.1 itself leads to a more precise result.

6.4. COROLLARY. *Let X be a random variable with density proportional to e^{-I}. Let e be a unit vector in $\mathbb{R}^d$. Under the assumptions of Theorem 6.1, the conditional distribution of $X-\lambda e$ given $X \in A+\lambda e$ converges weakly* to a point mass at $-rN(e) \in \partial A$ as λ tends to infinity.*

Proof. Let U be a neighborhood of $-rN(e)$. We can find a closed convex set B in U with smooth boundary and positive curvature, such that ∂B and ∂A coincide in a neighborhood of $-rN(e)$. Applying Theorem 6.1 twice, we see that

$$\lim_{\lambda\to\infty} \frac{\displaystyle\int_{B+\lambda e} e^{-I(x)}\mathrm{d}x}{\displaystyle\int_{A+\lambda e} e^{-I(x)}\mathrm{d}x} = 1\,.$$

Consequently, the conditional distribution of $X-\lambda e$ given $X \in A+\lambda e$ is asymptotically concentrated on $B \subset U$. Since U is an arbitrary small neighborhood of $-rN(e)$, the result follows. ∎

A slightly more involved proof would show that the conditional distribution of $X-t$ given $X \in A+t$ can be approximated by a point mass at $p_t - t$ as $|t|$ tends to infinity under the assumption of Theorem 6.1.

Notes

This chapter has three motivations. First it provides a simple example of applying of Theorem 5.1, and I hope it is of pedagogical interest. Second, translating a set away from the origin may be one of the most intuitive and natural ways to make it moving to infinity. Third, and this is more important, the Gaussian case has received some attention, due to statistical applications. In LeCam's theory of local asymptotic normality — see e.g., LeCam (1986) and LeCam and Yang (1990) — the asymptotic power of a test is given by the probability that a

noncentered Gaussian vector lies in a given domain. Thus, one issue is to calculate the probability that a centered Gaussian vector hits a translated set, typically an ellipsoid. The work of Breitung (1994), Breitung and Hohenbichler (1989), Breitung and Richter (1996) are most relevant here. The remarks following Proposition 6.3 somewhat enlighten the Gaussian case.

7. Homothetic sets, homogeneous I and Laplace's method

In this chapter we consider a set A_1 such that

$$\text{there exists a neighborhood of 0 not intersecting } A_1 \tag{7.1}$$

Equivalently, we could say that the complement of A_1 is a neighborhood of the origin. This assumption ensures that the sets $A_t = tA_1$ are moving to infinity as t tends to infinity. Assume furthermore that

$$I \text{ is a strictly convex, } \alpha\text{-positively homogeneous function,} \tag{7.2}$$

that is $I(tx) = t^\alpha I(x)$ for all nonnegative t, all x in $\mathbb{R}^d$, and some positive α. Under such assumptions, α must be strictly larger than 1 to ensure strict convexity. Setting $x = ty$, we see that

$$\int_{A_t} e^{-I(x)} \mathrm{d}x = t^d \int_{A_1} e^{-I(ty)} \mathrm{d}y = t^d \int_{A_1} e^{-t^\alpha I(y)} \mathrm{d}y \,.$$

The asymptotic decay of the last term in the equality is related to the Laplace method. When I has a unique minimum in A_1, not on the boundary of A_1, this type of integral has been well studied. However, here, $I(A_1)$ is achieved on the boundary of A_1, eventually on a k-dimensional submanifold of $\mathbb{R}^d$. A direct proof of an asymptotic equivalent of the right hand side, working out a multivariate Laplace method, is quite tractable. However, for purely pedagogical reasons, we will obtain an asymptotic equivalent of the right hand side of the above equality by using Theorem 5.1. This proof does not require more work than a direct one. The equality with the right hand side makes it easy to understand how Theorem 5.1 works. It also shows that Theorem 5.1 can be thought as a generalization of Laplace's method.

Consider $\mathcal{D}_{A_1} = A_1 \cap \Lambda_{I(A_1)}$ and assume that

$$\partial A_1,\ \mathcal{D}_{A_1} \text{ and } \Lambda_c \text{ are smooth — twice continuously differentiable — manifolds.} \tag{7.3}$$

Let k be the dimension of $\mathcal{D}_{A_1}$. We assume that A_1 separates from $\Lambda_{I(A_1)}$ with contact of order 1 exactly, and therefore,

$$\det G_{A_1}(p) \neq 0 \qquad \text{for any } p \in \mathcal{D}_{A_1}\,. \tag{7.4}$$

We also need to make sure that A_1 is not a thin $(d-1)$-dimensional layer against Λ_{A_1}. For instance we could assume that it is equal to the closure of its interior. Such an assumption is global. We can work with a much weaker local one. Roughly speaking, for p in $\mathcal{D}_{A_1}$, we need to be able to squeeze a ball in the intersection of A_1 with the forward image of a leaf $\pi_{A_1}^{-1}(p)$ through the normal flow. This guarantees some thickness near $\mathcal{D}_{A_1}$ along the section of A_1 orthogonal to $\mathcal{D}_{A_1}$. The exact assumption is that

$$\begin{gathered}\text{there exists a positive } \epsilon \text{ such that for all } p \text{ in } \mathcal{D}_{A_1} \text{ and}\\ \text{any unit vector } v \text{ in } T_p\pi_{A_1}^{-1}, \text{ any } s,\eta \text{ in } [\,0,\epsilon\,] \text{ the set}\\ A_1 \text{ contains } \psi\big[\,\exp_p(\eta v)\,,\tau_{A_1}\big(\exp_p(\eta v)\big)+s\,\big]\end{gathered} \tag{7.5}$$

The following is then a consequence of Theorem 5.1 and is a multivariate Laplace type approximation.

7.1. THEOREM. *Under (7.1)–(7.5), and if $\mathcal{D}_{A_1}$ is a base manifold for A_1, then*

$$\int_{A_t} e^{-I(x)}\mathrm{d}x \sim c_1 e^{-t^\alpha I(A_1)} t^{k-(\alpha-2)\frac{d-k}{2}-\frac{\alpha}{2}} \qquad \text{as } t\to\infty\,,$$

where

$$c_1 = (2\pi)^{\frac{d-k-1}{2}} \int_{\mathcal{D}_{A_1}} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_1}}}{|\mathrm{D}I|^{\frac{d-k+1}{2}}(\det G_{A_1})^{\frac{1}{2}}}\,.$$

As in Theorem 5.1, the asymptotic equivalent in Theorem 7.1 must be read with $\det G_{A_1} = 1$ if $\mathcal{D}_{A_1}$ is an open subset of $\Lambda_{I(A_1)}$ and $k = d-1$.

Before proving Theorem 7.1, notice first that for $\alpha = 2$, the polynomial term in t in the approximation has exponent $k-1$; it does not depend on the dimension d of the ambient space. More

importantly, no matter what d is, this exponent can be written as $k\frac{\alpha}{2}+d\frac{2-\alpha}{2}-\frac{\alpha}{2}$; it is an increasing function of k, as one should expect.

The proof of Proposition 4.1.1 shows that $G_{A_1}(p)=\Pi^\pi_{\Lambda_{I(A_1)},p}-\Pi^\pi_{\partial A_1,p}$ is the difference of the fundamental forms of $\Lambda_{I(A_1)}$ and ∂A_1 compressed to the direction orthogonal to $\mathcal{D}_{A_1}$.

During the proof of Theorem 7.1, we will make use of the following result, relating the large scale analysis of tA_1 to that of A_1 as far as the normal flow is concerned.

7.2. LEMMA *If I is α-homogeneous, then*
(i) $\psi(p,s)=t^{-1}\psi(tp,t^\alpha s)$, and
(ii) $\tau_{tA_1}(tp)=t^\alpha\tau_{A_1}(p)$.

Proof. To prove (i), write $\tilde\psi(p,s)=t^{-1}\psi(tp,t^\alpha s)$. Since $\mathrm{D}I$ is $(\alpha-1)$-homogeneous, Lemma 3.1.1 yields

$$\frac{\mathrm{d}}{\mathrm{d}s}\tilde\psi(p,s)=t^{\alpha-1}\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\bigl(\psi(tp,t^\alpha s)\bigr)=\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\bigl(\tilde\psi(p,s)\bigr)\,.$$

Thus, $\tilde\psi$ obeys the differential equation of Lemma 3.1.1. It equals ψ since $\tilde\psi(p,0)=p=\psi(p,0)$.

Assertion (ii) follows since $\tau_{A_1}(p)$ is the first positive time s such that $t\psi(p,s)$ is in tA_1, and

$$\psi\bigl(tp,t^\alpha\tau_{A_1}(p)\bigr)=t\psi\bigl(p,\tau_{A_1}(p)\bigr)$$

thanks to assertion (i). ∎

Proof of Theorem 7.1. Since all the assumptions used in Theorem 5.1 depend on $c_{A_t,M}$, we first need to guess its value, and then proceed. To this aim, we first need an estimate on the integral itself. We obtain it in evaluating the asymptotic equivalent given by Theorem 5.1.

It is natural to consider

$$\mathcal{D}_{A_t}=t\mathcal{D}_{A_1}=t\,\{\,x\in\partial A_1: I(x)=I(A_1)\,\}\,.$$

Using the homogeneity of I,

$$\mathrm{D}I(tp)=t^{\alpha-1}\mathrm{D}I(p)\,.$$

Let p be in $\mathcal{D}_{A_1}$, or equivalently, tp be in $\mathcal{D}_{A_t}$. To estimate $G_{A_t}(tp)$, the equality $(A_t,\Lambda_{I(A_t)})=(tA_1,t\Lambda_{I(A_1)})$ gives

$$G_{A_t}(tp)=t^{-1}G_{A_1}(p)$$

— the curvature tensors are rescaled by $1/t$; think of the sphere of radius t whose curvature is $1/t$ times that of a sphere of radius 1. We also have

$$\int f(p)\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_t}}(p) = t^k \int f(tq)\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_1}}(q)\,.$$

Consequently,

$$e^{-I(A_t)}(2\pi)^{\frac{d-k-1}{2}} \int_{\mathcal{D}_{A_t}} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_t}}}{|\mathrm{D}I|^{\frac{d-k+1}{2}}(\det G_{A_t})^{\frac{1}{2}}} = c_1 e^{-t^\alpha I(A_1)} t^{k-(\alpha-2)\frac{d-k}{2}-\frac{\alpha}{2}}$$

where c_1 is given in the statement of Theorem 7.1. Notice again that in a very few lines, Theorem 5.1 allows us to guess the result.

As pointed out in chapter 5, the larger $c_{A_t,M}$ is, the stronger the assumptions are. However, it is important to remember that all that we need is to find $c_{A_t,M}$ larger than $c^*_{A_t,M}$. So, consider a positive ϵ and let

$$c(t) = \Big(d\alpha - k + (\alpha-2)\frac{d-k}{2} + \frac{\alpha}{2} + \epsilon\Big)\log t\,.$$

From Proposition 2.1, we infer that

$$L\big(I(A_t)+c(t)\big) = L\big(t^\alpha I(A_1)+c(t)\big) = o(e^{t^\alpha I(A_1)} t^{k-(\alpha-2)\frac{d-k}{2}-\frac{\alpha}{2}}) \tag{7.6}$$

as t tends to infinity. Thus, $c^*_{A_t,M}$ is less than $c(t)$ for t large enough and any positive M. We can try to choose $c_{A_t,M}$ to be $c(t)$. In this case, we just proved that (5.4) is satisfied.

It should be noticed that our choice of $c(t)$ is very naive. We inverted asymptotically the function L and evaluated the inverse at the guessed asymptotic equivalent for $\int_{A_t} e^{-I(x)}\mathrm{d}x$. The addition of the term $\epsilon \log t$ in $c(t)$ is only to obtain (7.6).

We now proceed in checking all the assumptions needed to apply Theorem 5.1. We already chose a candidate for the dominating manifold,

$$\mathcal{D}_{A_t} = t\,\mathcal{D}_{A_1} = t\,\big\{\, x \in \partial A_1 : I(x) = I(A_1)\,\big\}\,.$$

We postpone the check of (5.1)–(5.2) to the end of the proof since it requires some discussion.

To describe $\underline{A}_{t,M}$, let

$$\tilde{A}_{1,t} = \{\, p \in \Lambda_{I(A_1)} : \tau_{A_1}(p) \leqslant c(t)/t^\alpha \,\} .$$

This choice ensures that

$$\underline{A}_{t,M} = \{\, p \in \Lambda_{I(A_t)} : \tau_{A_t}(p) \leqslant c(t) \,\} = t\tilde{A}_{1,t} .$$

Thus, we can look at $\underline{A}_{t,M}$ through its rescaled version $\tilde{A}_{1,t}$.

We check (5.3) through a rescaling. Let tq be in $\underline{A}_{t,M}$, and define $tp = \pi_{A_t}(tq)$. The point p is in $\mathcal{D}_{A_1}$. Lemma 7.2 implies

$$\tau_{A_t}(tq) - \tau_{A_t}(tp) = t^\alpha\big(\tau_{A_1}(q) - \tau_{A_1}(p)\big) .$$

Since q belongs to $\tilde{A}_{1,t}$, we have $\tau_{A_1}(q) \leqslant c(t)/t^\alpha$; in particular, $\tau_{A_1}(q)$ converges to 0 as t tends to infinity. Since τ_{A_1} is continuous on $\pi_{A_1}^{-1}(p)$ with a strict minimum on $\mathcal{D}_{A_1}$, we have $|q-p| = o(1)$ uniformly in q belonging to $\tilde{A}_{1,t}$ — recall $p = \pi_{A_1}(q)$. It then follows from Proposition 4.1.1 that

$$\tau_{A_1}(q) - \tau_{A_1}(p) = \frac{1}{2}\,|\mathrm{D}I(p)|\,\langle G_{A_1}(p)\exp_p^{-1}(q)\,,\exp_p^{-1}(q)\rangle\big(1+o(1)\big)$$

uniformly in $q \in \tilde{A}_{1,t}$. Consequently,

$$\tau_{A_t}(tq) - \tau_{A_t}(tp) = \frac{t^\alpha}{2}\,|\mathrm{D}I(p)|\,\big\langle G_{A_1}(p)\exp_p^{-1}(q)\,,\exp_p^{-1}(q)\big\rangle\big(1+o(1)\big)$$

as t tends to infinity, uniformly for tq in $\underline{A}_{t,M}$. This proves (5.3) here.

Notice that since $(\exp_p)_*(0) = \mathrm{Id}$, we also have the approximation

$$\tau_{A_1}(q) - \tau_{A_1}(p) = \frac{1}{2}\,|\mathrm{D}I(p)|\,\langle G_{A_1}(p)(q-p)\,,(q-p)\rangle\big(1+o(1)\big) ,$$

as q converges to p, which may look more familiar.

To verify (5.5), Lemma 7.1.2 shows that $\chi^F_{A_t}(tp) = t^\alpha\chi^F_{A_1}(p)$. Thus, it suffices to prove that $\inf_{p\in\tilde{A}_{1,t}}\chi^F_{A_1}(p)$ is uniformly bounded below by some positive number for t large enough. If q belongs to $\tilde{A}_{1,t}$, write $q = \exp_{\pi_{A_1}(q)}(\eta v)$ for $\eta = \mathrm{dist}\big(q,\pi_{A_1}(q)\big)$ and a unit vector v. As we have seen, η converges to 0 uniformly over q in $\tilde{A}_{1,t}$ as t tends to infinity. Therefore, (7.5) implies that for t large enough, $\chi^F_{A_1}(\cdot) \geqslant \epsilon$ over $\tilde{A}_{1,t}$. Thus, (5.5) holds.

Our choice of $c_{A_t,M} = c(t)$ ensures that (5.6) holds as well.

To check (5.7) is not much more complicated. Define

$$c = \sup\Big\{ \frac{\lambda_{\max}\big(\mathrm{D}^2 I(q)\big)^2}{\big|\mathrm{D}I(q)\big|^2} : q \in \Lambda_{I(A_1)} \Big\} .$$

We first notice that for any p in $\Lambda_{I(A_1)}$ and s positive, the definition of $K_{\max}$ in Proposition 4.2.4 and homogeneity of I imply

$$K_{\max}(tp,s) \leqslant \sup\Big\{ \frac{\lambda_{\max}\big(\mathrm{D}^2 I(tq)\big)^2}{|\mathrm{D}I(tq)|^2} : q \in \Lambda_{I(A_1)} \Big\} = \frac{c}{t^2} .$$

Identifying $T_p\Lambda_{I(p)}$ and $T_{tp}\Lambda_{I(tp)}$, the equality $\exp_{tp}(w) = t\exp_p(w/t)$ holds. It gives,

$$\begin{aligned} t_{0,M}(tp) &= \sup\Big\{ s : \inf_{v\in S_{T_{tp}\pi^{-1}_{A_t}(tp)}(0,1)} \tau_{tA_1}\big(\exp_{tp}(sv)\big) \leqslant c(t) \Big\} \\ &= t\sup\Big\{ s : \inf_{v\in S_{T_{tp}\pi^{-1}_{A_t}(tp)}(0,1)} \tau_{A_1}\big(\exp_{p}(sv)\big) \leqslant c(t)/t^\alpha \Big\} . \end{aligned}$$

Again, as $c(t)/t^\alpha$ converges to 0 as t tends to infinity, the requirement $\tau_{A_1}\big(\exp_p(sv)\big) \leqslant c(t)/t^\alpha$ forces sv to be $o(1)$ as t tends to infinity. Proposition 4.1.1 yields

$$t_{0,M}(tp) \leqslant t\sup\Big\{ s : \inf_{v\in S_{T_p\pi^{-1}_{A_1}(p)}(0,1)} \frac{1}{2}|\mathrm{D}I(p)|\langle G_{A_1}(p)sv\,,sv\rangle \leqslant \frac{2c(t)}{t^\alpha} \Big\}$$

for t large enough. Since $G_{A_1}(p)$ does not have null eigenvalues on $T_p\pi^{-1}_{A_1}(p)$ thanks to (7.2), $t_{0,M}(tp)$ is at most

$$\begin{aligned} t\sup\Big\{ s : s^2 \inf_{v\in S_{T_p\pi^{-1}_{A_1}(p)}(0,1)} \frac{1}{2}|\mathrm{D}I(p)|\langle G_{A_1}(p)v\,,v\rangle \leqslant \frac{2c(t)}{t^\alpha} \Big\} \\ = t^{1-\alpha/2}\sqrt{c(t)}\,O(1) \end{aligned}$$

as t tends to infinity. Consequently,

$$\sup_{tp\in\underline{A}_{t,M}} \sqrt{K_{\max}\big(tp,t_{0,M}(tp)\big)}t_{0,M}(tp) = t^{-\alpha/2}\sqrt{c(t)}\,O(1) = o(1)$$

as t tends to infinity, and (5.7) holds.

Following Remark 5.2, we will not check (5.8) but (5.16) instead. Adding subscripts to distinguish on which manifold we are working, we have

$$\mathcal{A}_{\pi^{-1}_{A_t}(tp)}(s,v) = t\,\mathcal{A}_{\pi^{-1}_{A_1}(p)}(s/t,v) ,$$

for all s in $\big[0, e_A(tp,v)\big)$ and all v in the unit sphere of $T_{tp}\pi_{A_t}^{-1}(tp)$. This unit sphere can be identified with that of $T_p\pi_{A_1}^{-1}(p)$. As we have seen, if $\exp_{tp}(sv)$ belongs to $\underline{A}_{t,M}$, then $t^{1-\alpha/2}s/t^\eta = o(1)$ for any positive η, as t tends to infinity, and uniformly in p belonging to $\mathcal{D}_{A_1}$. It follows from compactness of $\mathcal{D}_{A_1}$ and the classical expansion for $\mathcal{A}_{\pi_{A_1}^{-1}}(s,v)$ as s tends to 0 — see, e.g., Chavel (1996) —- that — recall $\alpha > 1$! —

$$\mathcal{A}_{\pi_{A_t}^{-1}(tp)}(s,v) = s\,\mathrm{Id}_{T_{tp}\pi_{A_t}^{-1}(tp)} + O(s^3/t^2) = s\,\mathrm{Id}_{T_{tp}\pi_{A_t}^{-1}(tp)} + o(1) \quad (7.7)$$

as t tends to infinity, and uniformly for p in $\mathcal{D}_{A_1}$. This implies (5.16).

Assumption (5.9) is easy to check. Since p belongs to $\mathcal{D}_{A_1}$ and q to $\tilde{A}_{1,t}\cap\pi_{A_1}^{-1}(p)$, we have

$$\frac{|\mathrm{D}I(tp)|}{|\mathrm{D}I(tq)|} = \frac{|\mathrm{D}I(p)|}{|\mathrm{D}I(q)|} = \frac{|\mathrm{D}I(tp)|}{\big|\mathrm{D}I\big(p+o(1)\big)\big|}$$

where the $o(1)$-term is uniform in q in the given range and p in $\mathcal{D}_{A_1}$.

Assumption (5.10) is trivial since

$$\frac{\|\mathrm{D}^2 I(tp)\|}{|\mathrm{D}I(tp)|^2} = t^{-\alpha}\frac{\|\mathrm{D}^2 I(p)\|}{|\mathrm{D}I(p)|^2}$$

by homogeneity of I.

Since $\tau_{A_1}(\cdot)$ vanishes on $\mathcal{D}_{A_1}$ here, (5.11) holds as well as (5.13).

We check (5.12) by rescaling. Indeed, for q in $\tilde{A}_{1,t}$, or equivalently, tq in $\underline{A}_{t,M}$, we have

$$\pi_{A_t}(tq) = t\pi_{A_1}(q)\,.$$

Identifying $T_{tq}\underline{A}_{t,M}$ and $T_q\tilde{A}_{1,t}$, we also have

$$\pi_{A_t,*}(tq)(v) = t\pi_{A_1,*}(q)(v/t)\,.$$

But $\big|q-\pi_{A_1}(q)\big|$ tends to 0 uniformly in $\tilde{A}_{1,t}$ as t tends to infinity. Since the differential $\pi_{A_1,*}\big(\pi_{A_1}(q)\big)$ is the orthogonal projection onto $T_p\mathcal{D}_{A_1}$ — this comes form the fact that $\pi_{A_1}(p) = p$ for all p in $\mathcal{D}_{A_1}$, and that $\pi_{A_1}^{-1}(p)$ is orthogonal to $\mathcal{D}_{A_1}$ — (5.12) follows.

We are left to check that in (5.1), $\mathcal{D}_{A_t}$ is indeed a dominating manifold for the set $A_t\cap\Gamma_{I(A)+c_{A_t},M}$ and that (5.2) holds. If $k=0$, i.e., $\mathcal{D}_{A_1}$ is made of a finite number of points, (5.1) is clear. Assumption (5.2) is then checked by rescaling, using the fact that $\tilde{A}_{1,t}$ shrinks around $\mathcal{D}_{A_1}$. Assume $k\geqslant 1$. Define

$$A_t^{(k)} = A_t\cap\big\{\,\psi\big(\exp_x(u),s\big) : x\in\mathcal{D}_{A_t}\,,\ u\in T_p\Lambda_{I(A_t)}\ominus T_p\mathcal{D}_{A_t}\,,\ s\geqslant 0\,\big\}\,.$$

Then, $\mathcal{D}_{A_t}$ is a dominating manifold for $A_t^{(k)} = tA_1^{(k)}$. From what we have done, we can apply Theorem 5.1 to obtain an asymptotic approximation for $\int_{A_t^{(k)}} e^{-I(x)}\mathrm{d}x$, and the result is nothing but the statement of Theorem 7.1. Instead of checking (5.1), it now suffices to prove that

$$\int_{A_t\setminus A_t^{(k)}} e^{-I(x)}\mathrm{d}x = o\Big(\int_{A_{t^{(k)}}} e^{-I(x)}\mathrm{d}x\Big) \qquad \text{as } t\to\infty\,.$$

To apply Theorem 5.1 to evaluate the integral in the left hand side of the above inequality, we need to find a dominating manifold for $A_t\setminus A_t^{(k)}$. Pick $\mathcal{D}_{A_1\setminus A_1^{(k)}}$ to be $\partial\mathcal{D}_{A_1}$. It is a dominating manifold, of dimension at most $k-1$ for the set

$$A_t^{(k-1)} = \big(A_t\setminus A_t^{(k)}\big)\bigcap\Big\{\,\psi(x,s) : x\in t\mathcal{D}_{A_1\setminus A_1^{(k)}}\,\Big\}\,.$$

Now, it is possible that $A_t\setminus A_t^{(k)}$ or $A_t^{(k-1)}$ do not have smooth boundaries. But since ∂A_1 is smooth, the set $A_t^{(k-1)}$ can be included in a set $A_t^{(k-1)\prime}$ say, with smooth boundary, and for which $t\mathcal{D}_{A_1\setminus A_1^{(k)}}$ is again a dominating manifold. This is done in parameterizing $\tau_{A_t^{(k-1)}}(p)$ for p in $t\mathcal{D}_{A_1\setminus A_1^{(k)}}$ so that $\tau_{A_t^{(k-1)\prime}}$ and $\tau_{A_t^{(k-1)}}$ coincide on the projection of $A_t^{(k-1)}$ on $\Lambda_{I(A_t)}$ for instance, and extending $\tau_{A_t^{(k-1)}}$ into a differentiable function on $\Lambda_{I(A_t)}$. This way, we need to estimate the integral over $A_t^{(k-1)\prime}$, for which we have a $(k-1)$-dimensional candidate for a dominating manifold. Iterating this process, we go down to a 0 dimensional dominating manifold, apply Theorem 5.1 in this case, and obtain the order of all the terms with dominating manifold of dimension between 0 and $k-1$. In particular, this implies

$$\begin{aligned}\int_{A_t\setminus A_t^{(k)}} e^{-I(x)}\mathrm{d}x &= O(1)e^{-I(A_t)}t^{(k-1)-(\alpha-2)\frac{d-(k-1)}{2}-\frac{\alpha}{2}}\\ &= o\Big(\int_{A_t^{(k)}} e^{-I(x)}\mathrm{d}x\Big)\,,\end{aligned}$$

and this proves Theorem 7.1. ∎

7.3. REMARK. It is essential to notice the following. In checking (5.3), (5.9), (5.12), the only feature we used besides homogeneity of I is that a point q in $\tilde{A}_{1,t}$ converges to $\pi_{A_1}(q)$ as t tends to infinity, uniformly in $\tilde{A}_{1,t}$. For (5.7) and (5.8) — actually (5.16) — we used

slightly more, namely that $\big|q-\pi_{A_1}(q)\big|=o(t^{-1/3})$ in order to obtain (7.7). The conclusion is that if I is homogeneous, all that we need to do to check the assumptions is to check that $\underline{A}_{t,M}/t$ concentrates to $\mathcal{D}_{A_t}/t$ at the rate $o(t^{-1/3})$. So, the work done in the proof of Theorem 7.1 may save us some effort in other applications.

7.4. REMARK. In many applications, the set A_1 is of the form $A_1=\big\{\,x : g(x)\geqslant 0\,\big\}$ for some smooth — twice continuously differentiable — function g, and $I(A)$ is achieved at boundary points x such that $g(x)=0$. In such cases, assumption (7.5) can be simplified into an analytical condition. To see this, define $q=\exp_p(\eta v)$ and $r=\psi\big(q,\tau_{A_1}(q)\big)$. Since

$$\psi\big(q,\tau_{A_1}(q)+s\big)=r+s\frac{\mathrm{D}I}{|\mathrm{D}I|^2}(r)+O(s^2)$$

thanks to Lemma 3.1.1, we have

$$\begin{aligned} g\Big(\psi\big(q,\tau_{A_1}(q)+s\big)\Big) &= g(r)+s\Big\langle \mathrm{D}g(r),\frac{\mathrm{D}I}{|\mathrm{D}I|^2}(r)\Big\rangle+O(s^2)\\ &= s\Big\langle \mathrm{D}g,\frac{\mathrm{D}I}{|\mathrm{D}I|^2}\Big\rangle(r)+O(s^2)\,. \end{aligned}$$

For this expression to be nonnegative for s nonnegative, it is enough to have

$$\langle \mathrm{D}g(r),\mathrm{D}I(r)\rangle>0\,.$$

Since I and g are smooth and q is close to p for small η, the condition

$$\inf_{p\in\mathcal{D}_{A_1}}\langle \mathrm{D}g(p),\mathrm{D}I(p)\rangle>0$$

is sufficient for (7.5) to hold. Since a point p in $\mathcal{D}_{A_1}$ minimizes I on A_1, the vectors $\mathrm{D}g(p)$ and $\mathrm{D}I(p)$ are positively proportional. The condition $\langle \mathrm{D}g(p),\mathrm{D}I(p)\rangle$ positive holds provided $\mathrm{D}g(p)$ is nonzero — notice that $\mathrm{D}I$ does not vanish for I is strictly convex. In other words, (7.5) is fulfilled as soon as g has no critical points on $\mathcal{D}_{A_1}$.

Applying Corollary 5.2, we can obtain results on conditional distributions. The following statement asserts that the probability measure with density propositional to $\mathrm{I}_{A_1}(x)e^{-I(tx)}$ converges weakly* as t tends to infinity to the probability measure absolutely continuous with respect to $\mathcal{M}_{\mathcal{D}_{A_1}}$, and density proportional to $|\mathrm{D}I|^{-(d-k+1)/2}(\det G_{A_1})^{-1/2}$.

7.5. THEOREM. *Under the assumptions of Theorem 7.1, if B is a set of continuity of $\mathcal{M}_{\mathcal{D}_{A_1}}$, then*

$$\lim_{t\to\infty} \frac{\displaystyle\int_{t(B\cap A_1)} e^{-I(x)}\mathrm{d}x}{\displaystyle\int_{tA_1} e^{-I(x)}\mathrm{d}x} = \frac{\displaystyle\int_{B\cap\mathcal{D}_{A_1}} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_1}}}{|\mathrm{D}I|^{(d-k+1)/2}(\det G_{A_1})^{1/2}}}{\displaystyle\int_{\mathcal{D}_{A_1}} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_1}}}{|\mathrm{D}I|^{(d-k+1)/2}(\det G_{A_1})^{1/2}}} .$$

Proof. We just need to check (5.18)–(5.20). Here, we consider $\lambda_{A_t} = t$. The measure in (5.18) rcan be rewritten as

$$\frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_1}}(q)}{|\mathrm{D}I|^{\frac{d-k+1}{2}}\big(\det G_{A_1}(q)\big)^{\frac12}} \Big/ \int_{A_1} \frac{\mathrm{d}\mathcal{M}_{\mathcal{D}_{A_1}}(q)}{|\mathrm{D}I|^{\frac{d-k+1}{2}}\big(\det G_{A_1}(q)\big)^{\frac12}}$$

and does not depend on t.

Since

$$\frac{c_{A_t,M}}{\lambda_{A_t}|\mathrm{D}I(tp)|} \leqslant 2d\alpha \frac{\log t}{t^\alpha |\mathrm{D}I(p)|} ,$$

assumption (5.19) is satisfied.

To check (5.20), use rescaling to obtain

$$\begin{aligned} \sup\Big\{ \frac{|tq-tp|}{\lambda_{A_t}} \,:\, tq\in\pi_{A_t}^{-1}(tp)\,,\ \tau_{A_t}(tq)\leqslant c(t)\Big\} \\ = \sup\big\{ |q-p| \,:\, q\in\pi_{A_1}^{-1}(p)\,,\ \tau_{A_1}(q)\leqslant c(t)/t^\alpha\,\big\} . \end{aligned}$$

Since $\tilde{A}_{1,t}$ shrinks around $\mathcal{D}_{A_1}$, assumption (5.20) holds true. Applying Corollary 5.2 yields the conclusion. ∎

Another way to formulate Theorem 7.5 is to say that the probability measures with density proportional to $\mathrm{I}_{A_1}(x)e^{-I(tx)}$ converge weakly* to the one with density proportional to $|DI|^{(d-k+1)/2}(\det G_{A_1})^{-1/2}$ with respect to the Riemannian measure on $\mathcal{D}_{A_1}$.

Notes

The notes of chapter 1 contain references on Laplace's method. Also very much related to this chapter is the work of Breitung (1994) in a Gaussian setting. Theorem 5.1 is related to Hwang (1980). The Laplace method in dimension larger than one with a dominating manifold of minimizing points is developed in Barbe and Broniatowski (200?), motivated by large deviation theory.

8. Quadratic forms of random vectors

In this chapter, we illustrate the use of Theorems 5.1 and 7.1 to deal with the following question. Consider a random vector $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$, and a real $d \times d$ matrix $C = (C_{i,j})_{1 \leqslant i,j \leqslant d}$. What is the decay of $P\{ \langle CX, X\rangle \geqslant t \}$ as t tends to infinity?

Of course, this decay depends on the distribution of X as well as on the matrix C. We will deal with two different types of distributions: symmetric Weibull- and Student-like. The Weibull-like tail will be handled through application of Theorem 7.1, while the Student-like one will be handled by a change of variable technique and Theorem 5.1.

8.1. An example with light tail distribution.

Consider a random vector $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$, having density $f = e^{-I}$ for some convex function I on $\mathbb{R}^d$. Assume moreover that I is α-positively homogeneous. Writing

$$A_t = \{ \, x \in \mathbb{R}^d : \langle Cx, x\rangle \geqslant t \, \} = \sqrt{t} A_1 \, ,$$

we see that

$$P\{ \, \langle CX, X\rangle \geqslant t \, \} = \int_{A_t} e^{-I(x)} \mathrm{d}x \, ,$$

and Theorem 7.1 is relevant here.

To be more specific, assume that the components X_i of X are independent and identically distributed, all with density

$$w_\alpha(x) = \frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)} \exp\Bigl(\frac{-|x|^\alpha}{\alpha} \Bigr) \, , \qquad x \in \mathbb{R} \, , \quad \alpha > 1 \, .$$

To apply Theorem 7.1 we need to describe the points of ∂A_1 at which

$$I(x) = \frac{1}{\alpha} \sum_{1 \leqslant i \leqslant d} |x_i|^\alpha$$

is minimum. Surprisingly, this problem seems quite difficult, and I have not been able to solve it in general. The result will rely on the following conjecture.

8.1.1. CONJECTURE. *If $\alpha \neq 2$ and $C + C^{\mathrm{T}}$ has no vanishing eigenvalue, then $\sum_{1\leqslant i\leqslant d}|x_i|^\alpha$ admits a finite number of minima in A_1; moreover $\det G_{A_1}$ is not null at these minima.*

Hence, if this conjecture is indeed true, $\mathcal{D}_{A_1}$ is of dimension $k = 0$ when $\alpha \neq 2$ and $C + C^{\mathrm{T}}$ has no degeneracy.

The application of Theorem 7.1 is then trivial. For $\alpha \neq 2$, we obtain

$$P\{\, \langle CX, X\rangle \geqslant t \,\} \sim \frac{\alpha^{d-(d/\alpha)}}{2^d\Gamma(1/\alpha)^d} c_1 e^{-t^{\alpha/2} I(A_1)} t^{-(\alpha-2)\frac{d}{4}-\frac{\alpha}{4}}$$

as t tends to infinity — recall that $A_t = \sqrt{t}A_1$ here and not tA_1 as in chapter 6 — where

$$c_1 = (2\pi)^{\frac{d-1}{2}} \sum_{x\in\mathcal{D}_{A_1}} |\mathrm{D}I(x)|^{-(d+1)/2} (\det G_{A_1})^{-1/2} \,.$$

The term in c_1 can be made more explicit. Indeed,

$$\mathrm{D}I(x) = \big(\mathrm{sign}(x_i)|x_i|^{\alpha-1}\big)_{1\leqslant i\leqslant d}$$

for $\alpha > 1$. Moreover, the remark following the statement of Theorem 7.1 asserts that G_{A_1} is obtained by the difference of the fundamental form of $\Lambda_{I(A_1)}$ and ∂A_1. Since $\mathrm{D}^2 I(x) = (\alpha-1)\mathrm{diag}\big(|x_i|^{\alpha-2}\big)$, we have

$$\Pi_{\Lambda_{I(A_1)}}(x) = \mathrm{Proj}_{T_x\Lambda_{I(A_1)}} \frac{\alpha-1}{|\mathrm{D}I(x)|} \mathrm{diag}\big(|x_i|^{\alpha-2}\big)\Big|_{T_x\Lambda_{I(A_1)}}$$

while

$$\Pi_{\partial A_1}(x) = \mathrm{Proj}_{T_x\partial A_1} \frac{C + C^{\mathrm{T}}}{|(C + C^{\mathrm{T}})x|}\Big|_{T_x\partial A_1} \,.$$

For $\alpha = 2$, the calculation can be done explicitly. Let λ be the largest eigenvalue of $C+C^{\mathrm{T}}$, and assume that λ is positive — otherwise $\langle Cx, x\rangle$ is nonpositive for any x and $P\{\, \langle CX, X\rangle \geqslant t \,\}$ is null for any positive t. Let

$$H = \{\, x : (C + C^{\mathrm{T}})x = \lambda x \,\}$$

be the eigenspace associated to the largest eigenvalue λ.

8.1.2. THEOREM. *Let k be $\dim H - 1$ and M be the compression of $\mathrm{Id} - \lambda^{-1}(C + C^{\mathrm{T}})$ to $H^\perp$. For $\alpha = 2$ and λ positive,*

$$P\{\, \langle CX, X\rangle \geqslant t\,\} \sim \frac{1}{\lambda^{(k-1)/2}\Gamma\big(\frac{k+1}{2}\big)(\det M)^{1/2}}\, e^{-t/\lambda}\, t^{(k-1)/2}$$

as t tends to infinity.

Proof. Assumptions (7.1), (7.2) and (7.3) hold. To check (7.4), we only need to calculate $\mathcal{D}_{A_1}$. We claim that $\mathcal{D}_{A_1}$ is the sphere of radius $\sqrt{2/\lambda}$ centered at the origin in H. Indeed, if $\langle Cx, x\rangle = 1$ then

$$2 = \langle (C + C^{\mathrm{T}})x, x\rangle \leqslant \lambda |x|^2 \,.$$

So $|x| \geqslant \sqrt{2/\lambda}$. On the other hand, if $(C+C^{\mathrm{T}})x = \lambda x$ and $|x|^2 = 2/\lambda$, then $\langle (C + C^{\mathrm{T}})x, x\rangle = \lambda|x|^2 = 2$, and then $\langle Cx, x\rangle = 1$.

Applying Theorem 7.1, we obtain,

$$\int_{A_t} e^{-|x|^2/2} \mathrm{d}x \sim c_1 e^{-tI(A_1)} t^{\frac{k-1}{2}} \qquad \text{as } t \to \infty \,,$$

and $I(A_1) = 1/\lambda$ from the preceding argument. To calculate the constant c_1, notice that in our case, $\mathrm{D}I = \mathrm{Id}$. Thus, $|\mathrm{D}I(x)| = \sqrt{2/\lambda}$ on $\mathcal{D}_{A_1}$.

To calculate G_{A_1}, observe that $T_x\Lambda_{I(x)} = \{\, x\,\}^\perp$, for the level lines of I are spheres. Thus, for x in $\mathcal{D}_{A_1}$, the second fundamental form of $\Lambda_{I(A_1)}$ at x is

$$\Pi_{\Lambda_{I(A_1)}}(x) = \frac{\mathrm{Proj}_{\{x\}^\perp} \mathrm{Id}\big|_{\{x\}^\perp}}{|\mathrm{D}I(x)|} = \sqrt{\frac{\lambda}{2}}\, \mathrm{Proj}_{\{x\}^\perp} \mathrm{Id}\big|_{\{x\}^\perp} \,.$$

On the other hand, the second fundamental form of ∂A_1 at some x in $\mathcal{D}_{A_1}$ is

$$\Pi_{\partial A_1}(x) = \frac{\mathrm{Proj}_{\{x\}^\perp}(C + C^{\mathrm{T}})\big|_{\{x\}^\perp}}{|(C + C^{\mathrm{T}})x|} = \frac{1}{\sqrt{2\lambda}} \mathrm{Proj}_{\{x\}^\perp}(C + C^{\mathrm{T}})\big|_{\{x\}^\perp} \,.$$

Clearly, since $\mathcal{D}_{A_1}$ is a sphere, its tangent space at x is $\{\, x\,\}^\perp \cap H$. It follows that for x belonging to $\mathcal{D}_{A_1}$,

$$T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1} = \{\, x\,\}^\perp \ominus \big(\{\, x\,\}^\perp \cap H\big) = \{\, x\,\}^\perp \ominus H = H^\perp$$

since x is in H as well. Thus, for x in $\mathcal{D}_{A_1}$,

$$G_{A_1}(x) = \mathrm{Proj}_{H^\perp}\sqrt{\frac{\lambda}{2}}\Big(\mathrm{Id} - \frac{(C+C^{\mathrm{T}})}{\lambda}\Big)\Big|_{H^\perp}.$$

This matrix does not depend on x. Therefore, putting all the pieces together,

$$c_1 = \frac{(2\pi)^{\frac{d-k-1}{2}}\mathrm{Vol}\big(S_H(0,\sqrt{2/\lambda})\big)}{(2/\lambda)^{\frac{d-k+1}{4}}(\lambda/2)^{\frac{d-k-1}{4}}\det\Big(\mathrm{Proj}_{H^\perp}\big(\mathrm{Id} - \frac{(C+C^{\mathrm{T}})}{\lambda}\big)\Big|_{H^\perp}\Big)^{\frac{1}{2}}}.$$

Using the classical fact

$$\mathrm{Vol}(S_{n-1}) = \frac{2\pi^{n/2}}{\Gamma\big(\frac{n}{2}\big)},$$

we obtain

$$c_1 = \frac{(2\pi)^{d/2}}{\lambda^{(k-1)/2}\Gamma\Big(\dfrac{k+1}{2}\Big)\det\Big(\big(\mathrm{Id} - \dfrac{(C+C^{\mathrm{T}})}{\lambda}\big)\Big|_{H^\perp}\Big)^{1/2}}.$$

The result follows after dividing c_1 by $(2\pi)^{d/2}$, the normalizing factor of the d-dimensional standard normal density. ∎

REMARK. It is interesting to notice the discontinuity in the polynomial term in t in Theorem 8.1.2. In (8.1.1), this term has degree $-(\alpha-2)\frac{d}{4} - \frac{\alpha}{4}$, which is strictly less than $-1/2$ for $\alpha \geqslant 2$, and equals $-1/2$ for $\alpha = 2$. For $\alpha = 2$, Theorem 8.1.2 gives a polynomial term of degree at most 0 since k is at least 1. Notice also that the map $C \mapsto k$ is not continuous for any standard topology on the set of matrices.

The determinant of M involved in Theorem 8.1.2 can be given more explicitly as a function of the matrix C. Write $\lambda_1 \leqslant \cdots \leqslant \lambda_d$ for the spectrum of $C + C^{\mathrm{T}}$. Since the dimension of H is $k+1$, we have $\lambda_{d-k} = \lambda_{d-k+1} = \ldots = \lambda_d$. Diagonalizing $C + C^{\mathrm{T}}$ and noticing that $H^\perp$ is invariant under $\mathrm{Id} + \lambda_d^{-1}(C + C^{\mathrm{T}})$, we have

$$\det M = \prod_{1\leqslant i\leqslant d-k-1}\Big(1 - \frac{\lambda_i}{\lambda_d}\Big).$$

Applying Theorem 7.5, we obtain also the following.

8.1.3. PROPOSITION. *For $\alpha = 2$ and λ positive, the conditional distribution of $X/\sqrt{t}$ given $\langle CX, X\rangle \geqslant t$ converges weakly* to the uniform distribution over the sphere centered at the origin of radius $\sqrt{2/\lambda}$ of H.*

Proof. Notice that $\mathcal{M}_{\mathcal{D}_{A_1}}$ is proportional to the uniform distribution on $S_H(0, \sqrt{2/\lambda})$ and that $|\mathrm{D}I|$ as well as $\det G_{A_1}$ are constant on $\mathcal{D}_{A_1}$. The result follows from the proof of Theorem 8.1.2. and Theorem 7.5. ■

8.2. An example with heavy tail distribution.

In this section, we consider a random vector $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$, with independent and identically distributed components, all having a Student-like distribution with parameter α. Thus, X_i has a density, and there exists a constant $K_{s,\alpha}$ such that

$$P\{\, X_i \leqslant -x \,\} \sim P\{\, X_i \geqslant x \,\} \sim \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{x^\alpha}$$

as x tends to infinity. Define

$$A_t = \big\{\, x \in \mathbb{R}^d : \langle Cx, x\rangle \geqslant t \,\big\} = \sqrt{t}A_1 \,.$$

Writing $s_\alpha(\cdot)$ for the density of a single X_i, the density of the vector X, given by $\prod_{1\leqslant i\leqslant d} s_\alpha(x_i)$. It is not log-concave. It is not even specified at all, except by an asymptotic equivalent! Thus, we cannot use Theorem 5.1 in a straightforward way to approximate

$$P\big\{\, \langle CX, X\rangle \geqslant t \,\big\} = \int_{A_t} \prod_{1\leqslant i\leqslant d} s_\alpha(x_i)\mathrm{d}x_i \,.$$

However, as pointed out in the introduction, chapter 1, we can make a change of variables, and then try to use Theorem 5.1. This will require all the power of Theorem 5.1, in particular the freedom on the set A that is allowed.

To state our first result, define

$$J_1 = \{\, j : C_{j,j} > 0 \,\} \,.$$

8.2.1. THEOREM. *Let X be a d-dimensional random vector with independent and identically distributed components having a Student-like distribution. Let C be a $d \times d$ matrix. If J_1 is not empty, then*

$$P\big\{\, \langle CX, X\rangle \geqslant t \,\big\} \sim K_{s,\alpha}\alpha^{(\alpha-1)/2}\frac{2}{t^{\alpha/2}}\sum_{j\in J_1} C_{j,j}^{\alpha/2} \qquad \textit{as } t \to \infty \,.$$

Proof. Let us first make a change of variable so that we will be able to apply Theorem 5.1.

Let $Y = (Y_1, \ldots, Y_d)$ be a random vector with centered normal distribution, with identity covariance matrix. Its density,

$$\frac{1}{(2\pi)^{d/2}} \exp\Big(-\frac{|y|^2}{2} \Big) , \qquad y \in \mathbb{R}^d ,$$

is log-concave. Let us write

$$\Phi(y) = \int_{-\infty}^{y} \frac{e^{-u^2}}{\sqrt{2\pi}} \mathrm{d}u , \qquad y \in \mathbb{R} ,$$

the normal cumulative distribution function. Similarly, denote by

$$S_\alpha(y) = \int_{-\infty}^{y} s_\alpha(u) \mathrm{d}u$$

the Student-like cumulative distribution function of each individual X_i. Writing

$$S_\alpha^{\leftarrow}(u) = \inf \{\, y : S_\alpha(y) \geqslant u \,\}$$

for the inverse function of S_α and analogously $\Phi^{\leftarrow}$ for the inverse function of Φ, we see that $S_\alpha^{\leftarrow} \circ \Phi(Y_i)$ has the same distribution as X_i.

NOTATION. *Let us agree that a function g defined on $\mathbb{R}$ is extended componentwise to $\mathbb{R}^d$, writing $g(x_1, \ldots, x_d)$ for $\big(g(x_1), \ldots, g(x_d)\big)$.*

It follows that X has the same distribution as $\langle C S_\alpha^{\leftarrow} \circ \Phi(Y) \, , S_\alpha^{\leftarrow} \circ \Phi(Y) \rangle$. In other words, defining

$$B_t = \{\, y \in \mathbb{R}^d : \langle C S_\alpha^{\leftarrow} \circ \Phi(y) \, , S_\alpha^{\leftarrow} \circ \Phi(y) \rangle \geqslant t \,\} = \Phi^{\leftarrow} \circ S_\alpha(A_t) ,$$

we have

$$P\{\, \langle CX, X \rangle \geqslant t \,\} = \int_{A_t} \prod_{1 \leqslant i \leqslant d} s_\alpha(x_i) \mathrm{d}x_i = \int_{B_t} \frac{e^{-|y|^2/2}}{(2\pi)^{d/2}} \mathrm{d}y . \qquad (8.2.1)$$

Since $S_\alpha^{\leftarrow} \circ \Phi$ is continuous and defined on the whole real line, we see that for any positive M and any t large enough, the ball centered at the origin and of radius M does not intersect B_t. Thus, B_t moves to infinity as t tends to infinity, and the right hand side of (8.2.1) is the integral of a log-concave function over a set moving to infinity as t tends to infinity. We can try to apply Theorem 5.1.

It should be noticed that we could make a change of variable leading to a different distribution than the standard Gaussian one. However, this one is rather convenient since its level sets and their geodesics are known explicitly.

The disadvantage of the change of variable is of course that the set B_t is more complicated than A_t. Nevertheless, whatever information is needed on B_t can be first read on A_t, and then pulled back to B_t. This fact is illustrated by Proposition 8.2.4 bellow, where we will calculate $\mathcal{D}_{B_t}$. This change of variable technique works mainly because $\Phi^{\leftarrow}\circ S_\alpha$ has an explicit and simple asymptotic equivalent.

To apply Theorem 5.1, let us define

$$I(y) = \frac{|y|^2}{2} + \log(2\pi)^{d/2}\,,$$

that is minus the logarithm of the Gaussian density. The function I is convex.

We will make use of the following elementary result in asymptotic analysis, whose proof can be found in appendix 1,

$$\big(\Phi^{\leftarrow}\circ S_\alpha(x)\big)^2 = 2\alpha\log x - \log\log x - 2\log(K_{s,\alpha}\alpha^{\alpha/2}) - 2\log(2\sqrt{\pi}) + o(1)$$

as x tends to infinity. It implies

$$\Phi^{\leftarrow}\circ S_\alpha(x) = \sqrt{2\alpha\log x} + o(1) \qquad \text{as } x\to\infty\,.$$

It is also convenient to introduce the canonical basis $e_1,\ldots,e_d$ of $\mathbb{R}^d$. For any j in J_1 and ϵ in $\{\,-1,1\,\}$, the point $p_{\epsilon,j,t} = \epsilon\sqrt{t/C_{j,j}}\,e_j$ belongs to ∂A_t. Thus, $q_{\epsilon,j,t} = \Phi^{\leftarrow}\circ S_\alpha(p_{\epsilon,j,t})$ belongs to ∂B_t. The following lemma gives a parameterization of ∂A_t and ∂B_t near $p_{\epsilon,j,t}$ and $q_{\epsilon,j,t}$. This describes these boundaries locally.

8.2.2. LEMMA. *The tangent space of the boundary ∂A_t at $p = p_{\epsilon,j,t}$ is $\{\,(C+C^{\mathrm T})p\,\}^{\perp}$. Near p, the boundary of A_t can be parametrized as*

$$p(v) = p_{\epsilon,j,t}(v) = \epsilon\sqrt{\frac{t}{C_{j,j}}}\Big[1 - \frac{1}{2t}\langle Cv,v\rangle + O\Big(\frac{\langle Cv,v\rangle}{t}\Big)^2\Big]e_j + v\,,$$

for v in $T_p\partial A_t$, and $|v| = o(\sqrt{t})$ as t tends to infinity.

The boundary ∂B_t near $q = q_{\epsilon,j,t} = \Phi^{\leftarrow}\circ S_\alpha(p)$ can be parametrized as

$$q(v) = q_{\epsilon,j,t}(v) = \epsilon\Bigg(\sqrt{\alpha\log\frac{t}{C_{j,j}}} - \frac{\log\log\sqrt{t}}{2\sqrt{\alpha\log t}} - \frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{\sqrt{\alpha\log t}} + o\Big(\frac{1}{\sqrt{\log t}}\Big)\Bigg)e_j + \sum_{\substack{1\leqslant i\leqslant d\\ i\neq j}} \Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$$

for v in $T_p\partial A_t$ and $|v| = o(\sqrt{t})$ as t tends to infinity.

Proof. The assertion on the tangent space of ∂A_t at p is plain since the differential of the map $x \mapsto \langle Cx, x\rangle$ at p is $(C + C^{\mathrm{T}})p$. Near the point $\sqrt{1/C_{i,i}}$, we can parameterize ∂A_1 by its tangent plane. This leads to the following parameterization of ∂A_t. Let $h(v)$ be such that

$$\tilde{p}(v) = \sqrt{t/C_{j,j}}\big(1 + h(v)\big)e_j + v \in \partial A_t$$

for all v in $T_p\partial A_t$ with $|v|$ not too large. This inclusion becomes

$$t = \langle C\tilde{p}(v), \tilde{p}(v)\rangle = t\big(1 + h(v)\big)^2 + \langle Cv, v\rangle\,. \qquad (8.2.2)$$

An approximation of h follows either by working out an asymptotic expansion for $h(\cdot)$ or using the following argument. For $|v| = o(\sqrt{t})$, (8.2.2) implies $h(v) = o(1)$ as t tends to infinity. Rewriting (8.2.2) as the quadratic equation in h,

$$0 = th(v)^2 + 2th(v) + \langle Cv, v\rangle\,,$$

we obtain

$$h(v) = -1 + \Big(1 - \frac{\langle Cv, v\rangle}{t}\Big)^{1/2} \qquad \text{as } t \to \infty\,.$$

This gives the asymptotic expansion for $p(v)$.

We then pull back the expression of $p(v)$ to parameterize ∂B_t by $q(v) = \Phi^{\leftarrow} \circ S_\alpha\big(p(v)\big)$. Notice first that

$$\log\left(\sqrt{\frac{t}{C_{j,j}}}\big(1 + O(t^{-1}|v|^2)\big)\right) = \frac{1}{2}\log\frac{t}{C_{j,j}} + O(t^{-1}|v|^2)\,.$$

In the range $|v| = o(\sqrt{t})$, the asymptotic expansion for $\Phi^{\leftarrow} \circ S_\alpha$ in Lemma A.1.5 gives

$$\begin{aligned}\epsilon\langle q(v), e_j\rangle &= \Phi^{\leftarrow} \circ S_\alpha\left(\sqrt{\frac{t}{C_{j,j}}}\big(1 + O(t^{-1}|v|^2)\big)\right) \\ &= \sqrt{\alpha\log\frac{t}{C_{j,j}}} - \frac{\log\log\sqrt{t}}{2\sqrt{\alpha\log t}} - \frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{\sqrt{\alpha\log t}} \\ &\qquad + o(\log t)^{-1/2}\,.\end{aligned}$$

On the other hand, for $i\neq j$,

$$\epsilon\langle q(v), e_i\rangle = \Phi^{\leftarrow}\circ S_\alpha(v_i) \ .$$

This proves Lemma 8.2.2. ∎

It is somewhat important for what follows to have some intuition on the shape of ∂B_t near $q_{\epsilon,j,t}$. This is precisely what the last assertion of Lemma 8.2.2 gives us. Recall that $p_{\epsilon,j,t}$ is collinear to e_j. As v varies in $T_{p_{\epsilon,j,t}}\partial A_t = \{\,(C+C^{\mathrm T})e_j\,\}^\perp$, the term $\sum_{1\leqslant i\leqslant d; i\neq j}\Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$ in the expression of $q_{\epsilon,j,t}(v)$ varies too. If the v_i's were allowed to vary independently, then $\sum_{1\leqslant i\leqslant d; i\neq j}\Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$ would describe the hyperplane $\{\,e_j\,\}^\perp$, and ∂B_t would be a hyperplane perpendicular to e_j, passing through $q_{\epsilon,j,t}$. This is not quite the case of course, but almost, provided we look at the right scale. This is the meaning of the next claim.

8.2.3. CLAIM. *For t large enough and j in J_1, the set*

$$\Big\{\sum_{\substack{1\leqslant i\leqslant d\\ i\neq j}}\Phi^{\leftarrow}\circ S_\alpha(v_i)e_i \,:\, v\perp (C+C^{\mathrm T})e_j\,,\, |v| = o(\sqrt t)\Big\}$$

is contained in

$$\Big\{\sum_{\substack{1\leqslant i\leqslant d\\ i\neq j}} w_ie_i \,:\, |w|\leqslant \frac12\sqrt{\alpha\log t}\Big\}\,.$$

Proof. One may argue as follows. Notice first that $\{\,(C+C^{\mathrm T})e_j\,\}^\perp$ does not contains e_j. Indeed, if this were the case, we would have $C_{j,j} = \langle (C+C^{\mathrm T})e_j\,,e_j\rangle = 0$, contradicting the fact that j belongs to J_1. Consequently, as v varies in $\{\,(C+C^{\mathrm T})e_j\,\}^\perp$, the vector $\sum_{1\leqslant i\leqslant d; i\neq j}\Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$ describes the space spanned by the e_i's for $1\leqslant i\leqslant d$ and $i\neq j$. Finally, if w is orthogonal to e_j and of norm less than $(1/2)\sqrt{\alpha\log t}$, then $w=\sum_{1\leqslant i\leqslant d; i\neq j}\Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$ for some v in $(C+C^{\mathrm T})e_j$. Furthermore, $\Phi^{\leftarrow}\circ S_\alpha(v_i)^2\leqslant \frac{\alpha}{2}\log t$. From Lemma A.1.5, we then infer $|v_i|\leqslant t^{3/8}$ for t large enough. The relation $v\perp (C+C^{\mathrm T})e_j$ forces then $|v_j|\leqslant O(t^{3/8})$, and so $|v| = o(\sqrt t)$. This proves our claim. ∎

We can now locate the interesting minima of I over ∂B_t. They will provide a good guess for a dominating manifold, as well as an estimation of $I(B_t)$.

8.2.4. PROPOSITION. *Assume that J_1 is nonempty. If y belongs to ∂B_t and $I(y) \leqslant I(B_t) + O(1)$ as t tends to infinity, then y is in a $O(1)$-neighborhood of a points $q_{\epsilon,j,t}$ for some j in J_1 and some ϵ in $\{\,-1,1\,\}$. Moreover, as t tends to infinity,*

$$I(q_{\epsilon,j,t}) = \frac{\alpha}{2}\log t - \frac{1}{2}\log\log\sqrt{t} - \log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) \\ + \log(2\pi)^{d/2} - \frac{\alpha}{2}\log C_{j,j} + o(1)\,.$$

Proof. By the very definition of I and B_t,

$$\begin{aligned} I(B_t) &- \log(2\pi)^{d/2} \\ &= \inf\big\{\,|y|^2/2 : \langle CS_\alpha^{\leftarrow}\circ\Phi(y)\,,S_\alpha^{\leftarrow}\circ\Phi(y)\rangle \geqslant t\,\big\} \\ &= \frac{1}{2}\inf\Big\{\sum_{1\leqslant i\leqslant d}\Phi^{\leftarrow}\circ S_\alpha(\lambda\sqrt{t}u_i)^2 : \langle Cu\,,u\rangle = 1\,,\,\lambda\geqslant 1\Big\} \\ &= \frac{1}{2}\inf\Big\{\sum_{1\leqslant i\leqslant d}\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u_i)^2 : \langle Cu\,,u\rangle = 1\Big\}\,, \end{aligned}$$

the second equality coming from the change of variable $y_i = \Phi^{\leftarrow}\circ S_\alpha(\lambda\sqrt{t}u_i)$; the last one comes from the fact that the function $\lambda\in[\,0,\infty)\mapsto\Phi^{\leftarrow}\circ S_\alpha(\lambda\sqrt{t}u_i)^2$ is increasing for t large enough and u_i fixed.

If s is positive and such that $\sqrt{t}s$ tends to infinity and $\log s/\log t$ tends to 0 as t tends to infinity, the asymptotic expansion for $(\Phi^{\leftarrow}\circ S_\alpha)^2$ in Lemma A.1.5. shows that

$$\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}s)^2 = \alpha\log t\big(1+o(1)\big) \qquad \text{as } t\to\infty\,. \tag{8.2.3}$$

If $\langle Cu\,,u\rangle = 1$, and r of the u_i's, say $u_1,\ldots,u_r$, are of order larger than $1/\log t$, i.e. $\min_{1\leqslant i\leqslant r}|u_i| \gg 1/\log t$ as t tends to infinity, then (8.2.3) yields

$$\sum_{1\leqslant i\leqslant d}\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u_i)^2 \geqslant \sum_{1\leqslant i\leqslant r}\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u_i)^2 \sim r\alpha\log t \tag{8.2.4}$$

as t tends to infinity. Hence, to minimize the left hand side of (8.2.4), we should have r as small as possible. But r must be at least 1, for $\langle Cu,u\rangle = 1$. Moreover $r=1$ can be achieved by considering j in J_1 and $u=\epsilon e_j/\sqrt{C_{j,j}}$ for some ϵ in $\{-1,1\}$. This leads us to look at the function I near $q_{\epsilon,j,t} = \Phi^{\leftarrow}\circ S_\alpha(\epsilon\sqrt{t}e_j/\sqrt{C_{j,j}})$. Furthermore, if $I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u)\big)$ is minimal, u must be on the boundary of a $O(\log t)^{-1}$-neighborhood of $\epsilon e_j/\sqrt{C_{j,j}}$ for some ϵ in $\{-1,1\}$ and j in J_1. Consequently, $\sqrt{t}u$ is in ∂A_t and in an $O(\sqrt{t}/\log t)=o(\sqrt{t})$-neighborhood of $p_{\epsilon,j,t}$. Therefore, when studying such a point, we can use the parameterization given in Lemma 8.2.2. This also leads us to look at the function $I\big(q_{\epsilon,j,t}(v)\big)$ for v in $T_{p_{\epsilon,j,t}}\partial A_t$ and $|v|=o(\sqrt{t})$.

Write $v_i=\langle v,e_i\rangle$ for the components of the vector v belonging to $T_{p_{\epsilon,j,t}}\partial A_t$. Using Lemma 8.2.2, we obtain

$$\begin{aligned} I\big(q_{\epsilon,j,t}(v)\big) &= \frac{\alpha}{2}\log\frac{t}{C_{j,j}} - \frac{1}{2}\log\log\sqrt{t} - \log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) \\ &\qquad + \frac{1}{2}\sum_{\substack{1\leqslant i\leqslant d\\ i\neq j}} \Phi^{\leftarrow}\circ S_\alpha(v_i)^2 + \log(2\pi)^{d/2} + o(1) \end{aligned} \tag{8.2.5}$$

as t tends to infinity, and uniformly in $|v|=o(\sqrt{t})$. Therefore, up to $o(1)$ as t tends to infinity, the function $v\mapsto I\big(q_{\epsilon,j,t}(v)\big)$ is minimum at 0, and its minimum value is $I(q_{\epsilon,j,t})$ as claimed. ∎

Notice that the proof of Proposition 8.2.4 gives actually a little bit more, and this will be useful. Indeed, if x is in ∂B_t and $I(x)=I(B_t)+o(\log\log t)$, then (8.2.3)–(8.2.4) and Lemma A.1.5 show that x is in a $o(\log\log t)$-neighborhood of some $q_{\epsilon,j,t}$. Indeed, we must have

$$\max_{\substack{1\leqslant i\leqslant d\\ i\neq j}} |\Phi^{\leftarrow}\circ S_\alpha(v_i)| = O(\log\log t)^{1/2}\,.$$

In view of Proposition 8.2.4 and its proof, we can start to apply Theorem 5.1 in calculating a few terms of the asymptotic formula. Indeed, define

$$\gamma_1 = \max_{1\leqslant j\leqslant d} C_{j,j}\,.$$

We have immediately

$$\begin{aligned} I(B_t) &= \frac{\alpha}{2}\log t - \frac{1}{2}\log\log\sqrt{t} - \log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) \\ &\qquad + \log(2\pi)^{d/2} - \frac{\alpha}{2}\log\gamma_1 + o(1) \end{aligned}$$

as t tends to infinity. Moreover, a candidate for the dominating manifold is

$$\{\, q_{\epsilon,j,t} : j \in J_1\,,\ \epsilon \in \{\,-1,1\,\}\,\}\,.$$

Unfortunately, this choice does not match with our definition of a dominating manifold. It is indeed required in the definition that it is also a base manifold, and as such belongs to $\Lambda_{I(B_t)}$. The expression $I(y) = (|y|^2/2) + \log(2\pi)^{d/2}$ and Proposition 8.2.4 shows that the points $q_{\epsilon,j,t}$, for j in J_1, cannot lie on the same sphere centered at the origin. But they almost do!

Let us denote by ρ_t the radius of the sphere $\Lambda_{I(B_t)}$. The expression for I shows that $\rho_t = \sqrt{I(B_t) + d\log(2\pi)}$. Define $r_{\epsilon,j,t} = \rho_t q_{\epsilon,j,t}/|q_{\epsilon,j,t}|$. We consider the candidate

$$\mathcal{D}_{B_t} = \{\, r_{\epsilon,j,t}\ :\ j \in J_1\,,\ \epsilon \in \{\,-1,1\,\}\,\}$$

for a dominating manifold. It will be clear after Lemma 8.2.7 that what we are really doing here is moving the points $q_{\epsilon,j,t}$ through the normal flow, until they reach the level line $\Lambda_{I(B_t)}$; this gives $r_{\epsilon,j,t}$ — somehow unfortunately for the clarity of the argument, but luckily for the calculation, this move along the normal flow and the Euclidean projection on the sphere coincide when working with the normal distribution.

Since $\mathcal{D}_{B_t}$ is of dimension 0, its Riemannian measure is a sum of point masses,

$$\mathcal{M}_{\mathcal{D}_{B_t}} = \sum_{\substack{j\in J_1\\ \epsilon\in\{-1,1\}}} \delta_{r_{\epsilon,j,t}}\,.$$

From Proposition 8.2.4 and the above expression for $I(B_t)$, we infer that for j in J_1,

$$\tau_{B_t}(r_{\epsilon,j,t}) = \frac{\alpha}{2}\log\frac{\gamma_1}{C_{j,j}} + o(1) \qquad \text{as } t\to\infty\,.$$

Since $\mathrm{D}I$ is the identity function and $I(y) = \dfrac{|y|^2}{2} + \log(2\pi)^{d/2}$, we also have

$$|\mathrm{D}I(r_{\epsilon,j,t})| = \rho_t = \sqrt{2I(B_t) + d\log(2\pi)} \sim \sqrt{\alpha\log t} \qquad \text{as } t\to\infty\,.$$

If we can apply Theorem 5.1, we obtain

$$
\begin{aligned}
&\int_{B_t} e^{-I(x)}\mathrm{d}x \\
&\quad\sim e^{-I(B_t)}(2\pi)^{(d-1)/2} \sum_{\substack{j\in J_1\\ \epsilon\in\{-1,1\}}} \frac{\exp\big(-\tau_{B_t}(q_{\epsilon,j})\big)}{|\mathrm{D}I(r_{\epsilon,j,t})|^{(d+1)/2}\big(\det G_{B_t}(r_{\epsilon,j,t})\big)^{1/2}} \\
&\quad\sim \frac{(\log t)^{(1-d)/4}}{t^{\alpha/2}} K_{s,\alpha}\alpha^{(2\alpha-1-d)/4} \sum_{\substack{j\in J_1\\ \epsilon\in\{1,1\}}} \frac{C_{j,j}^{\alpha/2}}{\big(\det G_{B_t}(r_{\epsilon,j,t})\big)^{1/2}} \qquad (8.2.6)
\end{aligned}
$$

as t tends to infinity. We are left with calculating $G_{B_t}(r_{\epsilon,j,t})$ and checking the assumptions of Theorem 5.1. In order to calculate $G_{B_t}(r_{\epsilon,j,t})$, we need to calculate τ_{B_t}, and ultimately the normal flow. This turns out to be particularly easy for the normal distribution.

8.2.5. LEMMA. *For the Gaussian distribution $\mathcal{N}(0,\mathrm{Id})$ on $\mathbb{R}^d$, the normal flow is given by $\psi(q,s)=\sqrt{1+\dfrac{2s}{|q|^2}}\,q$.*

Proof. The level lines Λ_c are spheres centered at the origin since I is a spherical function. Hence, $\psi(q,s)$ moves on a straight line through the origin as s varies, and $\psi(q,s)=a(s)q$ for some function $a(\cdot)$. We obtain a from the equation

$$
I(q)+s=\frac{|q|^2}{2}+\log(2\pi)^{d/2}+s=I\big(\psi(q,s)\big)=\frac{a(s)^2}{2}|q|^2+\log(2\pi)^{d/2}\,.
$$

That is, $a(s)=\sqrt{1+\dfrac{2s}{|q|^2}}\,.$ ∎

Since the exponential map on the level line $\Lambda_{I(B_t)}$ is involved in the definition of the curvature term G, we first recall its expression in the Gaussian case.

8.2.6. LEMMA. *If I is a spherical function, then for q in Λ_c, we have*
(i) $T_q\Lambda_c=\{\,q\,\}^\perp$,
(ii) $\exp_q(w)=\cos\Big(\dfrac{|w|}{|q|}\Big)q+\sin\Big(\dfrac{|w|}{|q|}\Big)|q|\dfrac{w}{|w|}$, *for all w in $T_q\Lambda_c$ with* $|w|\leqslant\pi|q|$.

Proof. Since Λ_c is a sphere centered at the origin, (i) follows. The geodesics on Λ_c are circles of maximal diameter. By cutting Λ_c along the plane determined by q and w, the expression of the maximal circle leaving q in the direction v, that is (ii), follows. ∎

Since we calculated $I(q_{\epsilon,j,t})$ in Proposition 8.2.4, it is easier to calculate $\tau_{B_t}\bigl(\exp_{q_{\epsilon,j,t}}(v)\bigr)$ than $\tau_{B_t}\bigl(\exp_{r_{\epsilon,j,t}}(v)\bigr)$. The following lemma will be instrumental in relating these quantities. It is specific to the Gaussian situation. Since $T_q\Lambda_{I(q)} = \{\,q\,\}^\perp$, we can identify $T_q\Lambda_{I(q)}$ and $T_{\lambda q}\Lambda_{I(\lambda q)}$ for any nonzero λ.

8.2.7. LEMMA. *For the standard Gaussian distribution, i.e., $I(x) = |x|^2/2 + \log(2\pi)^{d/2}$,*
(i) for any nonzero q, any positive λ and any w in $T_q\Lambda_{I(q)} \equiv T_{\lambda q}\Lambda_{I(\lambda q)}$, we have

$$\exp_{\lambda q}(w) = \lambda \exp_q(w/\lambda)\,;$$

(ii) moreover, for any set B, any positive λ, and any q in $\mathbb{R}^d$ such that the line segment between λq and q does not intersect B,

$$\tau_B(\lambda q) = \frac{|q|^2}{2}(1-\lambda^2) + \tau_B(q)\,.$$

Proof. (i) follows from Lemma 8.2.6 since

$$\exp_{\lambda q}(w) = \cos\Bigl(\frac{|w|}{\lambda|q|}\Bigr)\lambda q + \sin\Bigl(\frac{|w|}{\lambda|q|}\Bigr)|\lambda q|\frac{w}{|w|} = \lambda \exp_q\Bigl(\frac{w}{\lambda}\Bigr)\,.$$

To prove (ii), the condition that the segment between λq and q does not intersect B, the fact that the normal flow moves along straight lines through the origin, and the definition of τ_B imply $\psi\bigl(q,\tau_B(q)\bigr) = \psi\bigl(\lambda q,\tau_B(\lambda q)\bigr)$. The expression of the normal flow in Lemma 8.2.5 gives

$$\sqrt{1+2\frac{\tau_B(q)}{|q|^2}}\,q = \sqrt{1+2\frac{\tau_B(\lambda q)}{|\lambda q|^2}}\,\lambda q\,.$$

The formula for $\tau_B(\lambda q)$ follows. ∎

REMARK. The essence of Lemma 8.2.7 is to relate $\exp_{\psi(q,s)}\circ\psi_{s,*}$ and $\psi_s\circ\exp_q$. Both maps act on $T_q\Lambda_{I(q)}$. We can write one as the

other one composed with some transform of $T_q\Lambda_{I(q)}$. Lemma 8.2.7 makes this explicit in the Gaussian case.

We can obtain an approximation of τ_{B_t} provided its argument is not too far away from $q_{\epsilon,j,t}$ for some j in J_1 and some ϵ equal to -1 or $+1$. Since we made everything explicit up to $o(1)$, our approximation will not be good enough to check (5.3), but perfectly fine to check (5.17) — one may try to check (5.3) and hopefully will agree that (5.17) is a useful refinement. Hence, we are ready to calculate the curvature term G_{B_t}.

8.2.8. LEMMA. *In the range $|w| = o(\sqrt{\log t})$, we have*

$$\tau_{B_t}\big(\exp_{r_{\epsilon,j,t}}(w)\big) = \tilde\tau_{B_t}\big(\exp_{r_{\epsilon,j,t}}(w)\big) + o(1)\,,$$

with

$$\tilde\tau_{B_t}\big(\exp_{r_{\epsilon,j,t}}(w)\big) - \tilde\tau_{B_t}(r_{\epsilon,j,t}) = \frac{1}{2}\langle w,w\rangle\,.$$

Consequently,

$$G_{B_t}(r_{\epsilon,j,t}) \sim \frac{\mathrm{Id}_{\mathbb{R}^{d-1}}}{|\mathrm{D}I(q_{\epsilon,j})|} \sim \frac{\mathrm{Id}_{\mathbb{R}^{d-1}}}{\sqrt{\alpha\log t}} \qquad \textit{as } t\to\infty\,.$$

Proof. Let us first obtain an approximation for $\tau_{B_t}\big(\exp_{q_{\epsilon,j,t}}(w)\big)$. Lemma 8.2.2 shows that in the range $|v| = o(\sqrt{t})$, near $q_{\epsilon,j,t}$, the surface ∂B_t parametrized by $v\mapsto q_{\epsilon,j,t}(v)$, is given by the equation of the hyperplane $\langle x,e_j\rangle = |q_{\epsilon,j,t}|$, up to $o(\log t)^{-1/2}$. Consequently, for $q=q_{\epsilon,j,t}$, for w in $T_q\Lambda_{I(B_t)}$ and $s=\tau_{B_t}\big(\exp_q(w)\big)$, using Lemmas 8.2.5, 8.2.6, we obtain

$$|q| + o(\log t)^{-1/2} = \big\langle \psi\big(\exp_q(w),s\big),e_j\big\rangle = \sqrt{1+\frac{2s}{|q|^2}}\cos\Big(\frac{|w|}{|q|}\Big)|q|\,.$$

It follows that

$$\begin{aligned}\tau_{B_t}\big(\exp_q(w)\big) &= \frac{|q|^2}{2}\bigg[\bigg(\frac{|q|+o(\log t)^{-1/2}}{|q|\cos(|w|/|q|)}\bigg)^2 - 1\bigg]\\ &= \frac{|w|^2}{2} + o(1) \qquad \text{as } t\to\infty\,,\end{aligned}$$

in the range $|w| = o(|q|) = o(\log t)^{1/2}$. From Lemma 8.2.6, we deduce

$$\begin{aligned}\tau_{B_t}\bigl(\exp_{r_{\epsilon,j,t}}(w)\bigr) &= \tau_{B_t}\Bigl[\frac{\rho_t}{|q|}\exp_q\Bigl(w\frac{|q|}{\rho}\Bigr)\Bigr]\\ &= \frac{|q|^2}{2}\Bigl(1-\frac{\rho_t^2}{|q|^2}\Bigr)+\tau_{B_t}\Bigl[\exp_q\Bigl(w\frac{|q|}{\rho_t}\Bigr)\Bigr]\,.\end{aligned}$$

Therefore, since $\tau_{B_t}(q) = o(1)$,

$$\begin{aligned}\tau_{B_t}\bigl(\exp_{r_{\epsilon,j,t}}(w)\bigr)-\tau_{B_t}(r_{\epsilon,j,t}) &= \tau_{B_t}\Bigl[\exp_q\Bigl(w\frac{|q|}{\rho_t}\Bigr)\Bigr]+o(1)\\ &= \frac{|w|^2}{2}+o(1)\,.\end{aligned}$$

This is the result, setting

$$\tilde\tau_{B_t}\bigl(\exp_{r_{\epsilon,j,t}}(w)\bigr) = \frac{|q_{\epsilon,j,t}|^2-\rho_t^2}{2}+\frac{|w|^2}{2}\,.$$

The second statement follows since we proved the asymptotic equivalence $|\mathrm{D}I(r_{\epsilon,j,t})| \sim \sqrt{\alpha\log t}$. ∎

Combining Lemma 8.2.8 and result (8.2.6) yields the asymptotic equivalence given in Theorem 8.2.1. It remains to check the assumptions of Theorem 5.1.

Our choice of $\mathcal{D}_{B_t}$ as a discrete set ensures that (5.1) holds.

We now need a candidate for $c_{B_t,M}$. Let $c(t) = 2d\log\log t$. From Proposition 2.1 and our calculation of $I(B_t)$, we infer that

$$L\bigl(I(B_t)+c(t)\bigr) \leqslant \frac{1}{t^{\alpha/2}(\log t)^{d-\frac{1}{2}}}\,O(1) = o(t^{-\alpha/2}) \qquad \text{as } t\to\infty\,.$$

Thus, $c(t)$ is a good candidate for $c_{B_t,M}$, no matter what M is.

From the proof of Proposition 8.2.4 and Lemma 8.2.8, we infer that for any t large enough,

$$\begin{aligned}\underline{B}_{t,M} \subset \bigl\{\, &\exp_{r_{\epsilon,j,t}}(w) : |w| \leqslant \sqrt{5d\log\log t}\,;\\ &w\in T_{r_{\epsilon,j,t}}\Lambda_{I(B_t)}\,,\epsilon\in\{\,-1,1\,\}\,,\, j\in J_1\,\bigr\}\,.\end{aligned} \tag{8.2.7}$$

Since the level set $\Lambda_{I(B_t)}$ is a sphere of radius $\sqrt{2I(B_t)}\bigl(1+o(1)\bigr) \sim \sqrt{\alpha\log t}$ as t tends to infinity, its radius of injectivity is of order $\sqrt{\log t} \gg \sqrt{\log\log t}$, and assumption (5.2) holds.

Assumption (5.4) holds thanks to our choice of $c(t)$.

To verify (5.5), let r be a point in $\underline{B}_{t,M}$. The point $q = \psi\big(r, \tau_{B_t}(r)\big)$ is in the boundary ∂B_t. Since $\tau_{B_t}(r)$ is less than $c(t) = 2d \log\log t$, the proof of Proposition 8.2.4 and (8.2.5) show that $q = q_{\epsilon,j,t}(v)$ for some ϵ in $\{\,-1, 1\,\}$, some j in J_1 and

$$\sum_{\substack{1\leqslant i\leqslant d\\ i\neq j}} \Phi^{\leftarrow}\circ S_\alpha(v_i)^2 \leqslant 7d\log\log t\,.$$

Consequently, Lemma A.1.5 shows that $|v_i| \leqslant (\log t)^{2d/\alpha}$, for $1 \leqslant i \leqslant d$ with $i \neq j$. Since v is in $\{\,(C + C^{\mathrm{T}})e_i\,\}^{\perp}$, and this hyperplane does not contain e_j, we have $|v| = O(\log t)^{2d/\alpha}$. Notice that

$$\psi\big(r, \tau_{B_t}(r) + s\big) = \psi(q, s) = \sqrt{1 + \frac{2s}{|q|^2}}q\,,$$

and that $|q| \geqslant |r| = \rho_t$ tends to infinity with t. To prove that $\chi^F_{B_t} \geqslant \rho_t^2$ for instance — which is more than enough to guarantee (5.5) — it suffices to prove that λq is in B_t for any $1 \leqslant \lambda \leqslant 2$. This is plain from Lemma 8.2.2 and Claim 8.2.3.

Assumption (5.6) is plain.

To check (5.7), notice that for q in $\mathcal{D}_{B_t}$, the inequality $t_{0,M}(q) \leqslant \sqrt{5d\log\log t}$ holds thanks to (8.2.7). Furthermore, as $\mathrm{D}^2I(q)/|\mathrm{D}I(q)|$ equals $\mathrm{Id}/|q|$,

$$K_{\max}(q, t_0) \leqslant \sup\{\,|q|^{-2} : q \in \Lambda_{I(B_t)}\,\} \sim \frac{1}{2I(B_t)} \sim \frac{1}{\alpha\log t}$$

as t tends to infinity. Therefore, (5.7) holds.

Assumption (5.8) holds as well since $\pi_{B_t}^{-1}(p)$ is essentially a finite union of spherical caps, and the Ricci curvature of a sphere is positive.

Assumption (5.9) is trivially satisfied since Λ_c is a sphere. Thus, two points in Λ_c have equal norms.

It is no harder to verify (5.10), since

$$\frac{\|\mathrm{D}^2I(p)\|}{|\mathrm{D}I(p)|^2} = \frac{\|\mathrm{Id}\|}{|p|^2} = \frac{1}{|p|^2}\,.$$

To check assumption (5.11), again, we have $\|\mathrm{D}^2I\| = \|\mathrm{Id}\| = 1$. Moreover, if q belongs to $\underline{B}_{t,M} \subset \Lambda_{I(B_t)}$ and u is nonnegative,

$$\big|\mathrm{D}I\big(\psi_u(q)\big)\big|^2 \geqslant |\mathrm{D}I(q)|^2 \sim 2I(B_t) \sim \alpha\log t\,.$$

Consequently, for q in $\underline{B}_{t,M}$, that is $\tau_{B_t}(q)$ is less than $c(t)$, and for t large enough,

$$\int_0^{\tau_{B_t}(q)} \frac{\|\mathrm{D}^2 I\|}{|\mathrm{D}I|^2}\big(\psi_u(q)\big)\mathrm{d}u \leqslant \frac{2c(t)}{\alpha \log t} = \frac{4d \log\log t}{\alpha \log t}\,.$$

Since $\mathcal{D}_{B_t}$ is discrete, (5.12) holds automatically.

In conclusion, all the assumptions of Theorem 5.1 are satisfied. This proves Theorem 8.2.1. ∎

From the work done, we can easily infer the following conditional result.

8.2.9. THEOREM. *Under the assumption of Theorem 8.2.1, if J_1 is nonempty, the conditional distribution of the vector*

$$(\log\sqrt{t})^{-1}\big(\mathrm{sign}(X_i)\log|X_i|\big)_{1\leqslant i\leqslant d}$$

given $\langle CX, X\rangle \geqslant t$ converges weakly to*

$$\sum_{j\in J_1} \frac{C_{j,j}^{\alpha/2}}{\sum_{i\in J_1} C_{i,i}^{\alpha/2}} \frac{\delta_{-e_j}+\delta_{e_j}}{2}\,.$$

Proof. In order to apply Corollary 5.3, let us check its assumptions. Set $\lambda_{B_t} = \rho_t$. The numerator of the measure involved in (5.18) is

$$\sum_{\epsilon\in\{-1,1\}}\sum_{j\in J_1} \frac{\exp\big(-\tau_{B_t}(r_{\epsilon,j,t})\big)\delta_{r_{\epsilon,j,t}/\rho_t}}{|\mathrm{D}I(r_{\epsilon,j,t})|^{(d+1)/2}\big(\det G_{B_t}(r_{\epsilon,j,t})\big)^{1/2}}\,.$$

We already calculated

$$\tau_{B_t}(r_{\epsilon,j,t}) = I(q_{\epsilon,j,t}) - I(B_t) = \frac{\alpha}{2}\log\frac{\gamma_1}{C_{j,j}} + o(1) \qquad \text{as } t\to\infty\,.$$

Moreover, as $r_{\epsilon,j,t}$ is in the sphere $\Lambda_{I(B_t)}$ and $\mathrm{D}I = \mathrm{Id}$, we have $|\mathrm{D}I(r_{\epsilon,j,t})| = \rho_t$. Lemma 8.2.8 gives the value of $\det G_{B_t}(r_{\epsilon,j,t})$. Moreover, $r_{\epsilon,j,t}/\rho_t = q_{\epsilon,j,t}/|q_{\epsilon,j,t}| = \epsilon e_j$. Consequently, the measure in (5.18) is

$$\frac{\sum_{\epsilon\in\{-1,1\}}\sum_{j\in J_1} C_{j,j}^{\alpha/2}\big(1+o(1)\big)\delta_{\epsilon e_j}}{2\sum_{j\in J_1} C_{j,j}^{\alpha/2}\big(1+o(1)\big)}\,.$$

It certainly converges weakly* to

$$\nu = \frac{\sum_{j\in J_1} C_{j,j}^{\alpha/2}(\delta_{-e_j}+\delta_{e_j})}{2\sum_{j\in J_1} C_{j,j}^{\alpha/2}} .$$

Assumption (5.19) is trivial to verify since $\underline{B}_{t,M}$ is on the sphere $\Lambda_{I(B_t)}$ of radius ρ_t. Consequently, (5.19) becomes

$$\lim_{t\to\infty} \frac{2d\log\log t}{\rho_t^2} = \lim_{t\to\infty}\frac{2d\log\log t}{\alpha\log t} = 0 .$$

To check (5.20) is as simple. The inclusion (8.2.7) shows that if q belongs to $\underline{B}_{t,M}$; then the Riemannian distance on $\Lambda_{I(B_t)}$ between q and π_{B_t} is at most $\sqrt{5d\log\log t}$. Since $\Lambda_{I(B_t)}$ is a sphere of radius ρ_t, simple trigonometry shows that $|q-\pi_{B_t}(q)| \leqslant \rho_t \sin(\sqrt{5d\log\log t}/\rho_t)$. Since ρ_t is of order $\sqrt{\alpha\log t}$ as t tends to infinity, assumption (5.20) is fulfilled.

Applying Corollary 5.2, the distribution of $Y/\sqrt{\alpha\log t}$ given $Y\in B_t$ converges weakly* to ν. In other words, the distribution of $\Phi^{\leftarrow}\circ S_\alpha(X)/\sqrt{\alpha\log t}$ given $\langle CX, X\rangle \geqslant t$ converges weakly* to ν.

To rephrase this conclusion directly on X, we can use the Skorokhod (1956) representation theorem. It implies the existence of a random variable Y_t having the same distribution as Y given $Y\in B_t$, and a random variable Y_∞ having distribution ν such that $Y_t/\sqrt{\alpha\log t}$ converges almost surely to Y_∞. Thus, X given $\langle CX, X\rangle\geqslant t$ has the same distribution as

$$S_\alpha^{\leftarrow}\circ\Phi(Y_t\sqrt{\alpha\log t}) = S_\alpha^{\leftarrow}\circ\Phi\Big(Y_\infty\sqrt{\alpha\log t}\big(1+o(1)\big)\Big) .$$

Since $S_\alpha^{\leftarrow}\circ\Phi$ is ultimately sign preserving on $\mathbb{R}$ and Lemma A.1.6 yields

$$\begin{aligned}\log\Big|S_\alpha^{\leftarrow}\circ\Phi\Big(\epsilon e_j\sqrt{\alpha\log t}\big(1+o(1)\big)\Big)\Big| \\ &= e_j\log S_\alpha^{\leftarrow}\circ\Phi\Big(\sqrt{\alpha\log t}\big(1+o(1)\big)\Big) \\ &= e_j\frac{\log t}{2}\big(1+o(1)\big) ,\end{aligned}$$

the result follows ∎

A careful sharpening of all the estimates could certainly lead to more precise information on the conditional distribution of X

given $\langle CX, X\rangle \geqslant t$, and even an asymptotic expansion of this conditional distribution. We will not pursue in that direction for mainly two reasons: First, such calculation would be quite specific to this example. Second, we will see hereafter in this section that the kind of degeneracy at the limit — the limiting distribution is concentrated on a finite number of points — is not due to a bad rescaling but only to the fact that J_1 is nonempty.

It may happen that all the diagonal coefficients of the matrix C are nonpositive, that is $\max_{1\leqslant i\leqslant d} C_{i,i} \leqslant 0$. What is the analogue of Theorem 8.2.1 then? When γ_1 was positive, we could essentially set all the u_i's but one equal to 0 in order to optimize $I\bigl(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u)\bigr)$ — see the proof of Proposition 8.2.4 and inequality (8.2.4). When γ_1 is negative, we need to take at least two components u_i, u_j to be nonzero. Setting $p_{i,j} = \sqrt{t}\,u_ie_i + \sqrt{t}\,u_je_j$, the equation $p_{i,j}\in\partial A_t$ becomes

$$1 = u_i^2C_{i,i} + u_iu_j(C_{i,j}+C_{j,i}) + u_j^2C_{j,j}\,.$$

This equation admits a solution in u_i, u_j if and only if

$$(C_{i,j}+C_{j,i})^2 - 4C_{i,i}C_{j,j} > 0\,.$$

Consequently, if

$$J_2 = \bigl\{\,(i,j) : i\neq j\,,\ (C_{i,j}+C_{j,i})^2 - 4C_{i,j}C_{j,i} > 0\,\bigr\}$$

is nonempty, we can indeed consider only two nonzero components u_i, u_j with (i,j) in J_2.

How many components do we need to consider in general? To answer this question, it is more convenient to change the notation.

Let T be the set of all subsets of $\{\,1,2\ldots,d\,\}$. For a set $\mathcal{I} = \{\,i_1,\ldots,i_k\,\}$ with distinct elements, denote by $|\mathcal{I}| = k$ its cardinality. To $\mathcal{I}$, we associate the subspace $V_{\mathcal{I}} = \mathrm{span}\{\,e_{i_1},\ldots,e_{i_k}\,\}$ of dimension $|\mathcal{I}|$. To the matrix C, we associate

$$N(C) = \min\bigl\{\,|\mathcal{I}| : \mathcal{I}\in T\,,\ \exists u\in V_{\mathcal{I}}\,,\ \langle Cu,u\rangle > 0\,\bigr\}$$

and

$$J(C) = \bigl\{\,\mathcal{I}\in T : \exists u\in V_{\mathcal{I}}\,,\ \langle Cu,u\rangle > 0\,,\ |\mathcal{I}| = N(C)\,\bigr\}\,.$$

So, if $N(C)=1$, the set $J(C)$ is J_1. The integer $N(C)$ is the smallest cardinal of a set $\mathcal{I}$ such that the inequation $\langle Cu,u\rangle\geqslant 0$ has a solution in $V_{\mathcal{I}}\setminus\{\,0\,\}$. We exclude some degeneracy, assuming that

$$\text{for any } \mathcal{I} \text{ in } T \text{ of cardinal } |\mathcal{I}| < \mathcal{N}(\mathcal{C}), \text{ the matrix } C \text{ is negative on } V_{\mathcal{I}}, \tag{8.2.8}$$

that is $\langle Cu, u\rangle$ is negative for any nonzero vector u in $V_{\mathcal{I}}$. This typically prevents having γ_1 null, and the analogue when more components need to be considered. Notice that this nondegeneracy is typical with respect to the matrix C. In particular, the equation $\langle Cu, u\rangle = 1$ has a solution in any subspace $V_{\mathcal{I}}$ for $\mathcal{I}$ in $J(C)$, and has no solution in any subspace of the form $\mathrm{span}\{\, e_{\alpha_1}, \ldots, e_{\alpha_k} \,\}$ for any $k < N(C)$.

We keep the notation

$$A_t = \{\, x \in \mathbb{R}^d : \langle Cx, x\rangle = t \,\} = \sqrt{t} A_1 \,.$$

The sets

$$M_{\mathcal{I}} = \big\{\, m \in V_{\mathcal{I}} : \langle Cm \,, m\rangle = 1 \,\big\}\,, \qquad \mathcal{I} \in J(C)\,,$$

are $(|\mathcal{I}|-1)$-dimensional submanifolds of $\mathbb{R}^d$ and ∂A_1 as well. Indeed, it suffices to prove that 1 is a regular value of the map $x \in V_{\mathcal{I}} \mapsto \langle Cx, x\rangle$. The differential of this map at m is $(C+C^{\mathrm{T}})m$. If m belongs to $M_{\mathcal{I}}$,

$$\big\langle (C + C^{\mathrm{T}})m, m\big\rangle = 2\langle Cm, m\rangle = 2 \neq 0\,.$$

Consequently, $(C + C^{\mathrm{T}})m$ does not vanish, or, equivalently, the differential $(C + C^{\mathrm{T}})m$ is of full rank, and 1 is a regular value.

The result is then as follows.

8.2.10. THEOREM. *Let X be a d-dimensional random vector with independent and identically distributed components with Student-like distribution with parameter α. Let C be a $d\times d$ matrix and $N = N(C)$. For $\mathcal{I}$ in $J(C)$ and m in $M_{\mathcal{I}}$, denote $\tilde{G}(m)$ the compression of the diagonal matrix $\sum_{i\in\mathcal{I}} e_i \otimes e_i/|m_i|$ to $\big\{\, (C + C^{\mathrm{T}})m \,\big\}^{\perp} \cap V_{\mathcal{I}}$. Under (8.2.8), for $\alpha > 2/N$,*

$$P\{\, \langle CX, X\rangle \geqslant t \,\} \sim \frac{K_{s,\alpha}^N \alpha^{\alpha N/2}}{t^{\alpha N/2}\sqrt{\alpha N}} \sum_{\mathcal{I}\in J(C)} \int_{M_{\mathcal{I}}} \frac{\det \tilde{G}(m)}{\prod_{i\in\mathcal{I}} |m_i|^{\alpha}} \mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m)$$

as t tends to infinity.

REMARK. The assumption $\alpha > 2/N$ guarantees that the integral over $M_{\mathcal{I}}$ in the equivalence is finite. But we will see that the result is true whenever the integral over $M_{\mathcal{I}}$ is finite. It is not clear whether $\alpha > 2/N$ is required, though α too small makes the integral diverge. This can be seen in Lemma 8.2.18 below. If $N(C) = 1$, Theorem 8.2.10 is exactly Theorem 8.2.1. Clearly, in this

case, the term $K^N_{s,\alpha}\alpha^{\alpha N/2}t^{-\alpha N/2}/\sqrt{\alpha N}$ in Theorem 8.2.10 gives the term $K_{s,\alpha}\alpha^{(\alpha-1)/2}t^{-\alpha/2}$ in Theorem 8.2.1. For $N(C)=1$, we have $J(C)=J_1$, provided we identify $\{\,i\,\}$ and i. If $\mathcal{I}=\{\,i\,\}$ belongs to $J(C)$, then $V_{\mathcal{I}}=\mathbb{R}\,e_i$, and

$$M_{\mathcal{I}}=\big\{\,x\in\mathbb{R}\,e_i:\langle Cx,x\rangle=1\,\big\}=\{\,-C_{i,i}^{-1/2}e_i,C_{i,i}^{-1/2}e_i\,\}\,.$$

Thus, the Riemannian measure on $M_{\mathcal{I}}$ is

$$\mathcal{M}_{M_{\mathcal{I}}}=\delta_{-C_{i,i}^{-1/2}e_i}+\delta_{C_{i,i}^{-1/2}e_i}\,.$$

Moreover, for m in $M_{\mathcal{I}}$, the matrix $\tilde{G}(m)$ is the compression of $e_i\otimes e_i/m_i^2$ to $\big\{\,(C+C^{\mathrm{T}})m\,\big\}^{\perp}\cap V_{\mathcal{I}}$. But, in our case,

$$\big\{\,(C+C^{\mathrm{T}})m\,\big\}^{\perp}\cap V_{\mathcal{I}}=\emptyset\,,$$

because, if x is in $V_{\mathcal{I}}$, we have $x=se_i$ for some real number s, and

$$\big\langle(C+C^{\mathrm{T}})m\,,e_i\big\rangle=\pm C_{i,i}^{-1/2}\big\langle(C+C^{\mathrm{T}})e_i\,,e_i\big\rangle=\pm 2C_{i,i}^{1/2}\neq 0$$

— the last inequality holds since we assume $N(C)=1$ here and i in J_1, i.e., $\{\,i\,\}$ is in $J(C)$; we actually did the same work in the proof of claim 8.2.3. So, the term $\det\tilde{G}(m)$ has to be omitted, and we obtain

$$\sum_{\mathcal{I}\in J(C)}\int_{M_{\mathcal{I}}}\frac{\mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}}{\prod_{i\in\mathcal{I}}|m_i|^{\alpha}}=2\sum_{i\in J_1}C_{i,i}^{\alpha/2}\,,$$

as in Theorem 8.2.1.

Proof of Theorem 8.2.10. The proof is actually very similar to that of Theorem 8.2.1, except that the dominating manifold will no longer be a discrete set, and the parameterizations will be slightly more sophisticated.

As in the proof of Theorem 8.2.1, we denote by S_α the cumulative distribution function of X_i. We consider the two sets

$$A_t=\{\,x\in\mathbb{R}^d:\langle Cx,x\rangle\geqslant t\,\}=\sqrt{t}A_1\,,$$

and

$$B_t=\{\,y\in\mathbb{R}^d:\langle CS_\alpha^{\leftarrow}\circ\Phi(y)\,,S_\alpha^{\leftarrow}\circ\Phi(y)\rangle\geqslant t\,\}\,.$$

Notice that if $\mathcal{I}$ and $\mathcal{I}'$ are distinct and in $J(C)$, then $M_{\mathcal{I}}$ does not intersect $M_{\mathcal{I}'}$. Indeed, if x is in both $M_{\mathcal{I}}$ and $M_{\mathcal{I}'}$, then it is

in $V_{\mathcal{I}} \cap V_{\mathcal{I}'} = V_{\mathcal{I}\cap\mathcal{I}'}$, and moreover $\langle Cx, x\rangle = 1$. This contradicts the minimality of $N(C)$, since $\mathcal{I}$ different than $\mathcal{I}'$ implies $|\mathcal{I}\cap\mathcal{I}'| < N(C)$.

We first need a technical lemma, saying that whenever $\mathcal{I}$ is in $J(C)$, the manifold $M_{\mathcal{I}}$ stays away from any subspace $\{\, e_i \,\}^{\perp}$ with i in $\mathcal{I}$. We denote by $m_i = \langle m, e_i\rangle$ the components of the vector m belonging to $\mathbb{R}^d$.

8.2.11. LEMMA. *Under the assumption (8.2.8), there exists a positive ϵ_0 such that for all $\mathcal{I}$ in $J(C)$, all m in $M_{\mathcal{I}}$, and all i in $\mathcal{I}$, the inequality $|m_i| \geqslant \epsilon_0$ holds.*

Proof. Searching for a contradiction, assume that there exists $\mathcal{I}$ in $J(C)$, some index i in $\mathcal{I}$ and a sequence $m(n)$ in $M_{\mathcal{I}}$ such that $\lim_{n\to\infty} m_i(n) = 0$. Write

$$m(n) = m_i(n)e_i + s(n)v(n)$$

where $v(n)$ is a unit vector in $V_{\mathcal{I}\setminus\{\,i\,\}} = V_{\mathcal{I}} \ominus e_i\mathbb{R}$, and $s(n)$ is a real number. Dropping the index n for notational simplicity, the condition $m = m(n)$ belonging to $M_{\mathcal{I}}$ becomes

$$1 = \langle Cm\,, m\rangle = m_i^2\langle Ce_i\,, e_i\rangle + m_i s\langle (C + C^{\mathrm{T}})e_i\,, v\rangle + s^2\langle Cv\,, v\rangle\,.$$

Assumption (8.2.8) guarantees

$$\sup\bigl\{\, \langle Cv, v\rangle : v \in V_{\mathcal{I}\setminus\{\,i\,\}}\,,\ |v| = 1 \,\bigr\} < 0\,.$$

Thus, the above quadratic equation in s does not have any solution, since its discriminant is $m_i^2\langle (C + C^{\mathrm{T}})e_i, v\rangle - 4(m_i^2 C_{i,i} - 1)\langle Cv, v\rangle = 4\langle Cv, v\rangle + o(1)$, which is negative as n tends to infinity; this is a contradiction. ∎

In order to parameterize the boundaries ∂A_t and ∂B_t, we consider the normal bundle of the immersion $M_{\mathcal{I}} \subset \partial A_1$, namely,

$$\mathcal{N}_{\mathcal{I}} = \bigl\{\, (m, v) : m \in M_{\mathcal{I}}\,,\ v \in T_m\partial A_1 \ominus T_m M_{\mathcal{I}} \,\bigr\}\,.$$

The analogue of the parameterization $p_{\epsilon,j,t}(v)$ of ∂A_t in the proof of Theorem 8.2.1 is now a map defined on $o(\sqrt{t})$-sections of the normal bundle $\mathcal{N}_{\mathcal{I}}$.

It is convenient to introduce

$$Q(t) = \sqrt{\alpha\log t} - \frac{\log\log\sqrt{t}}{2\sqrt{\alpha\log t}} - \frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{\sqrt{\alpha\log t}}\,.$$

8.2.12. LEMMA. *Let m be in $M_{\mathcal{I}}$. The boundary ∂A_t near $\sqrt{t}m$ can be parameterized as*

$$p_{\mathcal{I},t}(m,v) = \sqrt{t}m\Bigl(1 - \frac{1}{2t}\langle Cv\,,v\rangle + o(1/t)\Bigr) + v\,,$$

$(m,v) \in \mathcal{N}_{\mathcal{I}}$, $|v| = o(\sqrt{t})$ as t tends to infinity. The boundary ∂B_t near $\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}m)$ can be parameterized as $q_{\mathcal{I},t}(m,v) + o(\log t)^{-1/2}$ where

$$q_{\mathcal{I},t}(m,v) = \sum_{i\in\mathcal{I}} \mathrm{sign}(m_i)\biggl(Q(t) + \frac{\sqrt{\alpha}\log|m_i|}{\sqrt{\log t}}\biggr)e_i + \sum_{\substack{1\leqslant i\leqslant d\\ i\notin\mathcal{I}}} \Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$$

and in the range $|v| = o(\sqrt{t})$, $\log|m_i| = o(\log t)^{1/2}$.

Proof. Since $\partial A_t = \sqrt{t}\partial A_1$ and ∂A_1 is a manifold, there exists a function h and some small positive ϵ such that for any $|v| \leqslant \epsilon\sqrt{t}$, any m in $M_{\mathcal{I}}$ and v in $T_m\partial A_1 \ominus T_m M_{\mathcal{I}}$,

$$p_{\mathcal{I},t}(m,v) = \sqrt{t}m\bigl(1 + h(v)\bigr) + v \in \partial A_t\,.$$

This equation can be rewritten as

$$t\langle Cm,m\rangle\bigl(1+h(v)\bigr)^2 + \sqrt{t}\bigl(1+h(v)\bigr)\langle (C+C^{\mathrm{T}})m\,,v\rangle + \langle Cv,v\rangle = t\,.$$

Since $\langle Cm,m\rangle = 1$ and $\langle (C+C^{\mathrm{T}})m,v\rangle = 0$ — recall that v belongs to $T_p\partial A_1 = \{\,(C+C^{\mathrm{T}})m\,\}^{\perp}$ — we obtain

$$0 = th(v)^2 + 2th(v) + \langle Cv\,,v\rangle$$

as in the proof of Lemma 8.2.2. Thus, $h(v) \sim -\langle Cv\,,v\rangle/(2t)$ as t tends to infinity, uniformly in $|v| = o(\sqrt{t})$. This gives the asymptotics for $p_{\mathcal{I},t}(m,v)$ in Lemma 8.2.12.

We pull back this parameterization to ∂B_t by introducing

$$\tilde q_{\mathcal{I},t}(m,v) = \Phi^{\leftarrow}\circ S_\alpha\bigl(p_{\mathcal{I},t}(m,v)\bigr)\,.$$

Lemma A.1.5 implies

$$\Phi^{\leftarrow}\circ S_\alpha\Bigl(\sqrt{t}m_i\bigl(1+|v|^2O(t^{-1})\bigr)+v_i\Bigr) = Q(t) + \sqrt{\alpha}\,\frac{\log|m_i|}{\sqrt{\log t}} + o(\log t)^{-1/2}$$

as t tends to infinity, uniformly in the range $|v| = o(\sqrt{t})$, $\log|m_i| = o(\log t)^{1/2}$ and $|m_i| \geqslant \epsilon_0 > 0$. Since Lemma 8.2.11 ensures that $|m_i|$

stays away from 0, the last condition, $|m_i| \geqslant \epsilon_0$, may be omitted in the statement of Lemma 8.2.12. ■

Define

$$R(t) = \frac{N(C)}{2}\alpha \log t - \frac{N(C)}{2} \log\log \sqrt{t} \\ - N(C) \log(K_{s,\alpha}\alpha^{\alpha/2} 2\sqrt{\pi}) + \log(2\pi)^{d/2} .$$

Recall that $I(x) = (|x|^2/2) + \log(2\pi)^{d/2}$. In the range $|v| = o(\sqrt{t})$ and $\log |m_i| = o(\log t)^{1/2}$, it follows from Lemma 8.2.12 that

$$\begin{aligned} I\big(q_{\mathcal{I},t}(m,v)\big) &= \frac{1}{2}\sum_{i\in\mathcal{I}} \Big(Q(t)^2 + 2Q(t)\frac{\sqrt{\alpha}\log|m_i|}{\sqrt{\log t}} + o(1)\Big) \\ &\qquad + \frac{1}{2}\sum_{\substack{1\leqslant i\leqslant d\\ i\notin\mathcal{I}}} \Phi^{\leftarrow}\circ S_\alpha(v_i)^2 + \log(2\pi)^{d/2} \\ &= R(t) + \alpha\sum_{i\in\mathcal{I}} \log|m_i| + \frac{1}{2}\sum_{\substack{1\leqslant i\leqslant d\\ i\notin I}} \Phi^{\leftarrow}\circ S_\alpha(v_i)^2 \\ &\qquad + o(1) \end{aligned} \tag{8.2.9}$$

as t tends to infinity, for all i in $\mathcal{I}$.

For a set $\mathcal{I}$ belonging to $J(C)$, define

$$\gamma_{\mathcal{I}} = \min\Big\{ \prod_{i\in\mathcal{I}} |m_i| : m \in M_{\mathcal{I}} \Big\} .$$

Furthermore, set

$$\gamma = \inf\{ \gamma_{\mathcal{I}} : \mathcal{I} \in J(C) \} .$$

The following result locates the points on ∂B_t where the function I is nearly minimal.

8.2.13. LEMMA. *We have*

$$I(B_t) = R(t) + \alpha\log\gamma + o(1) \qquad \textit{as } t\to\infty .$$

Moreover, there exists η in $(0,1/2)$ such that for any number positive M_1, any $M_2 > M_1/\alpha$, and any t large enough, the set

$$\partial B_t \cap \Gamma_{I(B_t)+M_1\log\log t}$$

is included in

$$\bigcup_{\mathcal{I}\in J(C)} \Big\{ q_{\mathcal{I},t}(m,v) : m \in M_{\mathcal{I}}\,,\ |m| \leqslant (\log t)^{M_2}\,, \\ |v| \leqslant t^{(1/2)-\eta}\,,\ |\mathrm{Proj}_{V_{\mathcal{I}}^{\perp}} v| \leqslant (\log t)^{M_2} \Big\}$$

where $\mathrm{Proj}_{V_{\mathcal{I}}^{\perp}}$ *is the projection onto* $V_{\mathcal{I}}^{\perp}$.

Proof. We first prove the second assertion of the lemma. By construction, $\Phi^{\leftarrow}\circ S_\alpha(\partial\sqrt{t}A_1) = \partial B_t$. Let u be a point in ∂A_1. Lemma A.1.5 implies that for any η in $(0,1/2)$ and provided t is large enough,

$$I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u)\big) \geqslant \alpha(1-\eta)\sum_{1\leqslant i\leqslant d} \mathrm{I}_{[t^{-\eta},\infty)}\big(|u_i|\big)\log\big(\sqrt{t}|u_i|\big)\,.$$

Consequently, if

$$I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}u)\big) \leqslant R(t) + M_1\log\log t + O(1)\,,$$

we must have, for t large enough,

$$(1+\eta)\frac{N(C)}{2}\alpha\log t \geqslant \alpha\,\sharp\big\{\,1\leqslant i\leqslant d : |u_i|\geqslant t^{-\eta}\,\big\}\log(t^{(1/2)-\eta})\,.$$

Taking η positive and small enough so that the integer part of $\frac{1+\eta}{1-2\eta}N(C)$ is $N(C)$, the previous inequality yields

$$N(C) \geqslant \sharp\big\{\,1\leqslant i\leqslant d : |u_i|\geqslant t^{-\eta}\,\big\}\,.$$

Then, since u is in ∂A_1, the minimality of $N(C)$ and Lemma 8.2.11 implies that we must have

$$N(C) = \sharp\big\{\,1\leqslant i\leqslant d : |u_i|\geqslant t^{-\eta}\,\big\}\,.$$

Thus, $\sqrt{t}u$ is in a $t^{(1/2)-\eta} = o(\sqrt{t})$-neighborhood of $\sqrt{t}\bigcup_{\mathcal{I}\in J(C)} M_{\mathcal{I}}$. Therefore, it can be written as $p_{\mathcal{I},t}(m,v)$ for some $\mathcal{I}$ in $J(C)$, some m in $M_{\mathcal{I}}$ and $|v| = o(t^{(1/2)-\eta})$.

Lemma 8.2.11 ensures that $m_i \geqslant \epsilon_0$ wherever i is in some set $\mathcal{I}$ of $J(C)$. Thus, for any positive ϵ and t large enough, Lemma A.1.5 implies

$$\begin{aligned} I\big(\Phi^{\leftarrow}\circ S_\alpha(p_{\mathcal{I},t})(m,v)\big) & \\ &\geqslant \frac{1}{2}\sum_{i\in\mathcal{I}}\Phi^{\leftarrow}\circ S_\alpha\big(\sqrt{t}|m_i|(1-\epsilon)\big)^2 \\ &\geqslant \alpha\sum_{i\in\mathcal{I}}\log\big(\sqrt{t}|m_i|\big) - \log\log\big(\sqrt{t}|m_i|\big) + o(1) \\ &= \Big(\frac{N(C)}{2}\alpha\log t + \alpha\sum_{i\in\mathcal{I}}\log|m_i|\Big)\big(1+o(1)\big) \qquad (8.2.10) \end{aligned}$$

Therefore, the inequality

$$I\Big(\Phi^{\leftarrow}\circ S_\alpha\big(p_{\mathcal{I},t}(m,v)\big)\Big) \leqslant R(t) + M_1\log\log t + O(1)$$

implies $\alpha\max_{i\in\mathcal{I}}\log|m_i| \leqslant M_1\log\log t\big(1+o(1)\big)$ for t large enough. So, for t large enough, $\log|m| \leqslant M_2\log\log t$ for M_2 larger than M_1/α. Notice that M_2 can be chosen independently of m in $M_{\mathcal{I}}$ since all the bounds are uniform in such m's. This proves the second assertion of Lemma 8.2.13, except for the restriction $|\mathrm{Proj}_{V_{\mathcal{I}}^\perp}v| \leqslant (\log t)^{M_2}$.

In the range obtained so far, $I\big(q_{\mathcal{I},t}(m,v)\big)$ has minimal value $R(t)$ up to $o(1)$ as t tends to infinity. Hence, $I(B_t)$ is as claimed. Finally, on this range, if $|\mathrm{Proj}_{V_{\mathcal{I}}^\perp}v| \geqslant (\log t)^{M_2}$, then at least one component v_j, for some j not in $\mathcal{I}$, is larger than $(\log t)^{M_2}/\sqrt{d}$. Consequently, as t tends to infinity,

$$\begin{aligned} I\big(\Phi^{\leftarrow}\circ S_\alpha(p_{\mathcal{I},t}(m,v)\big) &\geqslant R(t) + \alpha\log\gamma + \sum_{\substack{1\leqslant j\leqslant d\\ j\not\in I}} \frac{1}{2}\Phi^{\leftarrow}\circ S_\alpha(v_i)^2 + o(1) \\ &\geqslant R(t) + \alpha\log\gamma + \alpha M_2\log\log t\big(1+o(1)\big)\,, \end{aligned}$$

thanks to Lemma A.1.5. So, provided M_2 is large enough, the condition $q_{\mathcal{I},t}(m,v)$ belonging to $\Gamma_{I(B_t)+M_1\log\log t}$ imposes to have $|\mathrm{Proj}_{V_{\mathcal{I}}^\perp}v|$ less than $(\log t)^{M_2}$. ∎

As we did in the proof of Theorem 8.2.1, we can start to apply Theorem 5.1 in calculating the terms of the asymptotic formula.

How do we choose the dominating manifold?

Looking at the parameterization of ∂B_t by $q_{\mathcal{I},t}(m,v)$ in Lemma 8.2.12, we see that variations of $\prod_{i\in\mathcal{I}}|m_i|$ of order 1 yield small variations of order $1/\log t$ on $q_{\mathcal{I}}(m,v)$, i.e., in term of the Euclidean distance in $\mathbb{R}^d$, while, according to (8.2.9), they give variations of order 1 in term of I. This suggests that on the dominating manifold we should have $\prod_{i\in\mathcal{I}}|m_i|$ constant. Once m is fixed, (8.2.9) shows that fluctuations in space of order 1 in v yields fluctuations of order 1 on I as well. Keeping in mind that ∂B_t is of order $\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}) \sim \sqrt{\alpha\log t}$ as far as its size is concerned, we see that small fluctuations in v — on the scale of $\log t$ — brings sizable fluctuations of I. So, we should have v constant in the dominating manifold; and (8.2.9) suggests $v=0$. This leads us to consider the set

$$\Big\{\, m\in M_{\mathcal{I}} : \prod_{i\in\mathcal{I}}|m_i| = \gamma_{\mathcal{I}}\,\Big\}$$

of all points in $M_{\mathcal{I}}$ which minimize the product of their nonvanishing components. And then, one can try to choose the image of $\sqrt{t}$ times this set by $\Phi^{\leftarrow}\circ S_\alpha$ as the dominating manifold, i.e.,

$$\Big\{\, \Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}m) : m\in M_{\mathcal{I}}\,,\ \prod_{i\in\mathcal{I}}|m_i| = \gamma_{\mathcal{I}}\,\Big\}\,.$$

This is not quite right since it does not belong to $\Lambda_{I(B_t)}$, but a projection would fix this detail. For some reason that the author does not quite understand — can someone give an explanation? — this does not work and breaks down when looking for a quadratic approximation of τ_{B_t}. The right manifold to consider seems to be a projection of

$$\mathcal{D}'_{B_t} = \Big\{\, q_{\mathcal{I}}(m,0) : m\in M_{\mathcal{I}}\,,\ \max_{i\in\mathcal{I}}\log|m_i| \leqslant (\log t)^{1/4}\,,\ I\in J(C)\,\Big\}$$

on $\Lambda_{I(B_t)}$. The condition $\log|m_i| \leqslant (\log t)^{1/4}$ in the definition of $\mathcal{D}'_{B_t}$ guarantees that $\log|m_i| = o(\log t)^{1/2}$ and will allow us to use Lemma 8.2.12 and equality (8.2.9). Since $\Lambda_{I(B_t)}$ is a ball of radius $\rho_t = \sqrt{2\big(I(B_t)-\log(2\pi)^{d/2}\big)}$, this leads us to introduce

$$r_{\mathcal{I},t}(m,v) = \rho_t\frac{q_{\mathcal{I},t}(m,v)}{|q_{\mathcal{I},t}(m,v)|}\,,$$

and

$$\mathcal{D}_{B_t} = \Big\{\, r_{\mathcal{I},t}(m,0) : m\in M_{\mathcal{I}}\,,\ \max_{i\in\mathcal{I}}\log|m_i| \leqslant (\log t)^{1/4}\,,\ \mathcal{I}\in J(C)\,\Big\}\,.$$

Again, the projection on the sphere that we are doing is actually a mapping through the normal flow to the level set $\Lambda_{I(B_t)}$. In more general situations, we would define $r_{\mathcal{I},t}(m,v) = \psi_{B_t}\big(q_{\mathcal{I},t}(m,v),s\big)$ with $s = I(B_t) - I\big(q_{\mathcal{I},t}(m,0)\big)$.

Let us agree on the notation

$$q_{\mathcal{I},t}(m) = q_{\mathcal{I},t}(m,0) \qquad\text{and}\qquad r_{\mathcal{I},t}(m) = r_{\mathcal{I},t}(m,0)\,.$$

Define

$$p(m) = \sqrt{\alpha}\sum_{i\in\mathcal{I}}\operatorname{sign}(m_i)\log|m_i|e_i\,,\qquad m\in M_{\mathcal{I}}\,,\ \mathcal{I}\in J(C)\,.$$

For any $\mathcal{I}$ in $J(C)$ and any m in $M_{\mathcal{I}}$ with $\log|m| = o(\log t)^{1/2}$, we have

$$q_{\mathcal{I},t}(m) = Q(t)\sum_{i\in\mathcal{I}}\operatorname{sign}(m_i)e_i + \frac{p(m)}{\sqrt{\log t}}$$

as can be seen from Lemma 8.2.12. From Lemma 8.2.11, we infer that the vector $\sum_{i\in\mathcal{I}} \mathrm{sign}(m_i)e_i$ is constant on each connected component of $\bigcup_{\mathcal{I}\in J(C)} M_{\mathcal{I}}$. On each of these connected components, we can think of the set of all $q_{\mathcal{I},t}(m)$ — i.e., the connected components of $\mathcal{D}'_{B_t}$ — as a translation by a fixed vector of length $\sqrt{|\mathcal{I}|}Q(t)$ and a rescaling by $1/\sqrt{\log t}$ of the corresponding connected component of the set

$$\mathcal{D}_{\mathcal{I}} = \{\, p(m) : m \in M_{\mathcal{I}} \,\}\,, \qquad \mathcal{I} \in J(C)\,.$$

Following what we did in the proof of Theorem 8.2.1, we can start to apply Theorem 5.1. The value of $I(B_t)$ is available from Lemma 8.2.13.

For m in $M_{\mathcal{I}}$, the differential $p_*(m)$ is the restriction to $T_m M_{\mathcal{I}} = \{\, (C+C^{\mathrm{T}})m \,\}^{\perp} \cap V_{\mathcal{I}}$ of the matrix

$$\sqrt{\alpha}\,\mathrm{diag}(1/|m_i|)_{i\in\mathcal{I}} = \sqrt{\alpha}\sum_{i\in\mathcal{I}} e_i \otimes e_i/|m_i|\,.$$

The change of variable $m \leftrightarrow q_{\mathcal{I},t}(m)$ gives

$$\mathrm{d}\mathcal{M}_{\mathcal{D}'_{B_t}} = \det\bigl(p_*^{\mathrm{T}} p_*\bigr)^{1/2} \mathrm{d}\mathcal{M}_{\cup_{\mathcal{I}\in J(C)} M_I} (\log t)^{(1-N(C))/2}\,,$$

provided we integrate on the restricted range $\log|m| = o(\log t)^{1/2}$. Since $r_{\mathcal{I},t}(m)$ is reasonably close to $q_{\mathcal{I},t}(m)$, we should be able to approximate the measure $\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}}$ by $\mathrm{d}\mathcal{M}_{\mathcal{D}'_{B_t}}$ when applying Theorem 5.1.

From equation (8.2.9), we infer that

$$\tau_{B_t}\bigl(q_{\mathcal{I},t}(m)\bigr) = \alpha\sum_{i\in\mathcal{I}} \log|m_i| - \alpha\log\gamma + o(1) \qquad \text{as } t\to\infty\,.$$

To obtain the needed quadratic approximation for τ_{B_t} along the geodesic leaves orthogonal to $T_r\mathcal{D}_{B_t}$, we first need to determine $T_r\Lambda_{I(B_t)} \ominus T_r\mathcal{D}_{B_t}$. What should it be? Let us argue informally. Later, we will make the argument rigorous. First, we should be able to replace $T_r\mathcal{D}_{B_t}$ by $T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t}$ for some $\mathcal{I}$ in $J(C)$ and m in $M_{\mathcal{I}}$. Since $T_{q_{\mathcal{I},t}(m)}\Lambda_{I(B_t)}$ is the orthocomplement of $q_{\mathcal{I},t}(m)\mathbb{R}$, we should expect $T_{q_{\mathcal{I},t}(m)}\Lambda_{I(B_t)} \ominus T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t}$ to be roughly $\{\, q_{\mathcal{I},t}(m)\,\}^{\perp} \ominus T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t}$. Given the expression for $q_{\mathcal{I},t}(m)$, we have

$$\begin{aligned} T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t} &= T_{p(m)}\mathcal{D}_{\mathcal{I}} = p_*(m) T_m M_{\mathcal{I}} \\ &= \Bigl\{ \sum_{i\in\mathcal{I}} \frac{v_i}{|m_i|} e_i : v \in V_{\mathcal{I}} \ominus (C+C^{\mathrm{T}})m\mathbb{R} \Bigr\} \subset V_{\mathcal{I}}\,. \end{aligned}$$

Intuitively, our idea of projecting $\mathcal{D}'_{B_t}$ on $\Lambda_{I(B_t)}$ to obtain $\mathcal{D}_{B_t}$ will work well if the projection does not create a singularity or reduce the dimension; in other words, if $\mathcal{D}'_{B_t}$ is transverse to the direction of $q_{\mathcal{I},t}(m) = Q(t)\sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)e_i\bigl(1+o(1)\bigr)$, or, roughly, if $\sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)e_i$ does not belong to $T_{p(m)}\mathcal{D}_{\mathcal{I}}$ for all m in $M_{\mathcal{I}}$. Since this latter vector is in $M_{\mathcal{I}}$, our calculation of $T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t}$ allows to rewrite this condition as

$$\Bigl\langle \sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)|m_i|e_i\,,(C+C^{\mathrm{T}})m\Bigr\rangle \neq 0\,.$$

It holds since $\sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)|m_i|e_i = m$ and $\langle (C+C^{\mathrm{T}})m,m\rangle \geqslant 1$. In this case, since $q_{\mathcal{I},t}(m)\mathbb{R}$ is approximately $\sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)e_i\mathbb{R}\subset V_{\mathcal{I}}$, the subspace $V_{\mathcal{I}}$ is almost spanned by $q_{\mathcal{I},t}(m)\mathbb{R}$ and $T_{p(m)}\mathcal{D}_{\mathcal{I}}$. So, approximately,

$$\bigl\{\,q_{\mathcal{I},t}(m)\,\bigr\}^{\perp}\ominus T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t}\approx V_{\mathcal{I}}^{\perp}\,.$$

Then, the expression of $q_{\mathcal{I},t}$ in Lemma 8.2.12 shows that in $V_{\mathcal{I}}^{\perp}+q_{\mathcal{I},t}(m)\mathbb{R}$, the boundary ∂B_t is approximately the $\bigl(d-N(C)\bigr)$-dimensional affine space $V_{\mathcal{I}}^{\perp}+q_{\mathcal{I},t}(m,0)$ — consider $\Phi^{\leftarrow}\circ S_\alpha(v_i)$ as a new coordinate, say, w_i. But then, working on $V_{\mathcal{I}}^{\perp}+q_{\mathcal{I},t}(m)\mathbb{R}$, the very same argument as in the proof of Theorem 8.2.1 shows that $\tau_{B_t}\bigl(\exp_{q_{\mathcal{I},t}(m)}(w)\bigr)$ is approximately $|w|^2/2$ for w in the normal bundle. So, we should have

$$G_{B_t}\bigl(q_{\mathcal{I},t}(m)\bigr)\approx\frac{\mathrm{Id}_{V_{\mathcal{I}}^{\perp}}}{|q_{\mathcal{I},t}(m)|}=\frac{\mathrm{Id}_{\mathbb{R}^{d-N(C)}}}{Q(t)\sqrt{N(C)}}\,.$$

Thus, as t tends to infinity,

$$\begin{aligned}\det G_{B_t}\bigl(q_{\mathcal{I},t}(m)\bigr)^{-1/2}&\sim Q(t)^{(d-N(C))/2}N(C)^{(d-N(C))/4}\\&\sim\bigl(\alpha N(C)\log t\bigr)^{(d-N(C))/4}\,.\end{aligned}$$

We can then put all the pieces together, find a way to drop the restriction $\log|m_i|\leqslant(\log t)^{1/4}$ in the range of integration on $\mathcal{D}_{B_t}$, and use the formula given in Theorem 5.1 to obtain the — hypothetical — approximation

$$\begin{aligned}P\bigl\{\,\langle CX,X\rangle\geqslant t\,\bigr\}\sim\frac{K_{s,\alpha}^{N(C)}\alpha^{\alpha N(C)/2}}{\sqrt{N\alpha}\,t^{\alpha N(C)/2}}\times&\\ \int_{\cup_{\mathcal{I}\in J(C)}M_{\mathcal{I}}}\frac{\det\bigl(p_*(m)^{\mathrm{T}}p_*(m)\bigr)^{1/2}}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\,\mathrm{d}\mathcal{M}_{\cup_{\mathcal{I}\in J(C)}M_{\mathcal{I}}}(m)\,,&\end{aligned}$$

which is the result.

Let us now work out the proper arguments.

We first determine our candidate for c_{B_t}. Let

$$c(t)=\Big(d+\frac{N(C)}{2}+1\Big)\log\log t\,.$$

From Proposition 2.1 and Lemma 8.2.13, we infer that

$$\begin{aligned}L\big(I(B_t)+c(t)\big)&=O(1)e^{-I(B_t)}e^{-c(t)}I(B_t)^d\\&=t^{-\alpha N(C)/2}(\log t)^{d+(N(C)/2)}(\log t)^{-d-(N(C)/2)-1}O(1)\\&=o\big(t^{-\alpha N(C)/2}\big)\,.\end{aligned}\qquad(8.2.11)$$

Thus, $c(t)$ is a good candidate for an upper bound of c_{B_t} — notice again that $c(t)$ is of order $\log I(B_t)$. Many assumptions in Theorem 5.1 deal with the behavior of τ_{B_t} near the dominating manifold. Our goal is now to more or less calculate the value of τ_{B_t} where we need it.

The following lemma will give us the normal bundle of $\mathcal{D}_{B_t}$ immersed in $\Lambda_{I(B_t)}$.

8.2.14. LEMMA. *Let $\mathcal{I}$ be in $J(C)$. Let m be a point in $M_{\mathcal{I}}$ with $\max_{i\in\mathcal{I}}\log|m_i|=o(\log t)^{1/2}$. Then, $T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}+r_{\mathcal{I},t}(m)\mathbb{R}=V_{\mathcal{I}}$. Consequently,*

$$T_{r_{\mathcal{I},t}(m)}\Lambda_{I(B_t)}\ominus T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}=V_{\mathcal{I}}^{\perp}\,.$$

Proof. The map $u\in\mathbb{R}^d\setminus\{\,0\,\}\mapsto u/|u|\in S_{d-1}$ has differential $|u|^{-1}\mathrm{Proj}_{u^\perp}$ at u. Consequently,

$$\begin{aligned}&T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}\\&\qquad=\frac{\rho_t}{|q_{\mathcal{I},t}(m)|}\mathrm{Proj}_{q_{\mathcal{I},t}(m)^\perp}T_{q_{\mathcal{I},t}(m)}\mathcal{D}'_{B_t}\\&\qquad=\mathrm{Proj}_{q_{\mathcal{I},t}(m)^\perp}\{\,q_{\mathcal{I},t,*}(m)v:v\in T_mM_{\mathcal{I}}\,\}\\&\qquad=\mathrm{Proj}_{q_{\mathcal{I},t}(m)^\perp}\Big\{\sum_{i\in\mathcal{I}}\frac{v_i}{|m_i|}e_i:\langle(C+C^{\mathrm T})m\,,v\rangle=0\Big\}\subset V_{\mathcal{I}}\,.\end{aligned}$$

Consequently, $T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}$ is of dimension

$$\dim\big(\{\,(C+C^{\mathrm T})m\,\}^\perp\cap V_{\mathcal{I}}\big)=\dim V_{\mathcal{I}}-1=N(C)-1$$

and $T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}+q_{\mathcal{I},t}(m)\mathbb{R}=V_{\mathcal{I}}$. Since $r_{\mathcal{I},t}(m)$ is collinear to $q_{\mathcal{I},t}(m)$, this gives the first statement of the lemma.

The last statement of Lemma 8.2.14 follows from the fact that the tangent space of $\Lambda_{I(B_t)}$ at $r_{\mathcal{I},t}(m)$ is orthogonal to $q_{\mathcal{I},t}(m)$ since the level sets are spheres. ■

As we did in the proof of Theorem 8.2.1 with Lemma 8.2.7, we now relate τ_{B_t} on $\mathcal{D}_{B_t}$ to τ_{B_t} on $\mathcal{D}'_{B_t}$. Since $\mathcal{D}'_{B_t}$ is somewhat more explicit than $\mathcal{D}_{B_t}$, and better parametrized, this will make further calculations easier. Using Lemma 8.2.7, we see that for any w orthogonal to $r_{t,I}(m)$,

$$\begin{aligned}\tau_{B_t}\big(\exp_{r_{\mathcal{I},t}(m)}(w)\big) &= \tau_{B_t}\Big[\frac{\rho_t}{|q_{\mathcal{I},t}(m)|}\exp_{q_{\mathcal{I},t}(m)}\Big(w\frac{|q_{\mathcal{I},t}(m)|}{\rho_t}\Big)\Big]\\ &= \frac{|q_{\mathcal{I},t}(m)|^2}{2}\Big(1-\frac{\rho_t^2}{|q_{\mathcal{I},t}(m)|^2}\Big)\\ &\qquad + \tau_{B_t}\Big[\exp_{q_{\mathcal{I},t}(m)}\Big(w\frac{|q_{\mathcal{I},t}(m)|}{\rho_t}\Big)\Big]\end{aligned}\tag{8.2.12}$$

In particular, since $\exp_q(0)=q$,

$$\begin{aligned}&\tau_{B_t}\big(\exp_{r_{\mathcal{I},t}(m)}(w)\big)-\tau_{B_t}\big(r_{\mathcal{I},t}(m)\big)\\ &\qquad =\tau_{B_t}\Big[\exp_{q_{\mathcal{I},t}(m)}\Big(w\frac{|q_{\mathcal{I},t}(m)|}{\rho_t}\Big)\Big]-\tau_{B_t}\big(q_{\mathcal{I},t}(m)\big)\\ &\qquad =\tau_{B_t}\Big[\exp_{q_{\mathcal{I},t}(m)}\Big(w\frac{|q_{\mathcal{I},t}(m)|}{\rho_t}\Big)\Big]+o(1)\,,\end{aligned}$$

the last equality coming from the parameterization of ∂B_t in Lemma 8.2.12. We can now prove the analogue of Lemma 8.2.8.

8.2.15. LEMMA. *For m in $\mathcal{D}_{B_t}$ and w in $V_{\mathcal{I}}^{\perp}=T_{r_{\mathcal{I},t}(m)}\Lambda_{I(B_t)}\ominus T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}$ in the range $|w|=o(\sqrt{\log t})$, we have*

$$\tau_{B_t}\big(\exp_{r_{\mathcal{I},t}(m)}(w)\big)-\tau_{B_t}\big(r_{\mathcal{I},t}(m)\big)=\frac{|w|^2}{2}+o(1)\qquad \textit{as } t\to\infty\,.$$

Consequently,

$$G_{B_t}\big(r_{\mathcal{I},t}(m)\big)=\frac{\mathrm{Id}_{\mathbb{R}^{d-N(C)}}}{\big|\mathrm{D}I\big(r_{\mathcal{I},t}(m)\big)\big|}\sim\frac{\mathrm{Id}_{\mathbb{R}^{d-N(C)}}}{\sqrt{N(C)\alpha\log t}}\qquad \textit{as } t\to\infty\,.$$

Proof. Let w be a vector orthogonal to $V_{\mathcal{I}}$. From the calculation preceding Lemma 8.2.15, we see that it is enough to evaluate $s = \tau_{B_t}(z)$ for $z = \exp_{q_{\mathcal{I},t}(m)}\big(w|q_{\mathcal{I},t}(m)|/\rho_t\big)$. By definition, $\psi(z,s)$ is in the boundary of B_t. Lemma 8.2.12 shows that on directions orthogonal to $V_{\mathcal{I}}$, the boundary ∂B_t behaves like a $\big(d - N(C)\big)$-dimensional linear subspace — at least in the range $w = \Phi^{\leftarrow} \circ S_\alpha(v)$ with $v = o(\sqrt{t})$, that is $|w| = o(\log t)^{1/2}$. So, looking on the components on $V_{\mathcal{I}}$, we must have for (m,v) in $\mathcal{N}$,

$$\big\langle \psi(z,s)\,, e_i\big\rangle = \big\langle q_{\mathcal{I},t}(m,v), e_i\big\rangle + o(\log t)^{-1/2}\,, \qquad i \in \mathcal{I}\,.$$

To evaluate the left hand side term of the above equality, we use

$$\begin{aligned}\psi(z,s)&\\ &= \sqrt{1+\frac{2s}{|q_{\mathcal{I},t}(m)|^2}}\Big[\cos\Big(\frac{|w|}{\rho_t}\Big)q_{\mathcal{I},t}(m) + \sin\Big(\frac{|w|}{\rho_t}\Big)\big|q_{\mathcal{I},t}(m)\big|\frac{w}{|w|}\Big]\,.\end{aligned}$$

To evaluate the right hand side, we have

$$\langle q_{\mathcal{I},t}(m,v)\,, e_i\rangle = \langle q_{\mathcal{I},t}(m,0)\,, e_i\rangle + o(\log t)^{-1/2}\,.$$

Then, we obtain

$$s = \frac{|q_{\mathcal{I},t}(m,s)|^2}{2}\Big(\frac{1+o(\log t)^{-1}}{\cos^2\big(|w|/\rho_t\big)} - 1\Big) = \frac{|w|^2}{2} + o(1)$$

in the range $|w| = o(\rho_t)$, i.e., $|w| = o(\log t)^{1/2}$ as announced. The matrix $G_{B_t}\big(r_{\mathcal{I},t}(m)\big)$ in the statement of Lemma 8.2.15 is that corresponding to the assumption (5.17) as weakened at the end of Remark 5.2. ∎

We are now in position to verify that the assumptions of Theorem 5.1 hold.

It will be helpful to keep in mind two orders of magnitudes. Since $\mathcal{D}_{B_t}$ is contained in $\Lambda_{I(B_t)} = S_{d-1}(0,\rho_t)$, points in $\mathcal{D}_{B_t}$ are of order ρ_t, that is of order $\sqrt{\log t}$. On the other hand, Lemmas 8.2.12, 8.2.13 and A.1.5 show that any point of $\partial B_t \cap \Gamma_{I(B_t)+c(t)}$ is at a distance at most $O(\log\log t)$ of a point $q_{\mathcal{I},t}(m)$. Indeed, if $|\mathrm{Proj}_{V_{\mathcal{I}}^\perp} v| \leqslant (\log t)^{M_2}$, the component of $q_{\mathcal{I},t}(m,v)$ on $V_{\mathcal{I}}^\perp$ is $\sum_{1\leqslant i\leqslant d\,;\, i\not\in\mathcal{I}} \Phi^{\leftarrow}\circ S_\alpha(v_i)e_i$, which is of order $O(\log\log t)$ thanks to Lemma A.1.5. Since the projection onto the sphere $\Lambda_{I(B_t)}$ is a Lipschitz function when acting on $\Gamma^{\mathrm{c}}_{I(B_t)}$, the points in $\underline{B}_{t,M}$ are also at a distance $O(\log\log t)$ of $\mathcal{D}_{B_t}$.

Assumption (5.1) holds trivially since $k = \dim M_{\mathcal{I}} = N-1$.

To check (5.2) amounts to proving that any point s on $\Lambda_{I(B_t)}$, in a $O(\log\log t)$-neighborhood of $\mathcal{D}_{B_t}$, can be written in a unique way as $\exp_{r_{\mathcal{I},t}(m)}(w)$ for some $\mathcal{I}$ in $J(C)$, some m in $M_{\mathcal{I}}$ and

$$w \in T_{r_{\mathcal{I},t}(m)}\Lambda_{I(B_t)} \ominus T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t} \equiv V_{\mathcal{I}}^{\perp}\,.$$

Lemma 8.2.11 implies that such a point s is actually in an $O(\log\log t)$-neighborhood of a unique set $\mathcal{D}_{B_t} \cap V_{\mathcal{I}} = \big\{\, r_{\mathcal{I},t}(m) : m \in M_{\mathcal{I}} \,\big\}$ for some $\mathcal{I}$ in $J(C)$. For w belonging to $T_{r_{\mathcal{I},t}(m)}\Lambda_{I(B_t)} \ominus T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}$, we have

$$\mathrm{Proj}_{V_{\mathcal{I}}^{\perp}} \exp_{r_{\mathcal{I},t}(m)}(w) = \cos\Big(\frac{|w|}{\rho_t}\Big) r_{\mathcal{I},t}(m)\,.$$

Consequently, the component of s on $V_{\mathcal{I}}$ is in a one-to-one relation with $\cos\big(|w|/\rho_t\big) r_{\mathcal{I},t}(m)$. This last point being a positive multiple of $r_{\mathcal{I},t}(m) \in S_{d-1}(0,\rho_t)$, it identifies $r_{\mathcal{I},t}(m)$ and consequently m. Looking at the component of s in $V_{\mathcal{I}}^{\perp}$, we can then calculate w in a unique way, and (5.2) holds.

Our choice of $c(t)$ satisfying (8.2.11) will imply (5.4) ultimately, while (5.6) is plain.

Assumption (5.5) can be verified exactly as in the proof of Theorem 8.2.1. Indeed, Lemmas 8.2.12 and 8.2.13 show that $\partial B_t \cap \Gamma_{I(B_t)+c(t)}$ can be approximated by a ruled hypersurface based on $\mathcal{D}'_{B_t}$ where the generators are $(d-N)$-dimensional Euclidean balls of radius $O(\log\log t)$.

Assumption (5.7) is verified exactly in the same way as in the proof of Theorem 8.2.1. The bound $t_{0,M}(p) = O(\log\log t)$ follows from Lemma 8.2.15 — or from the discussion at the beginning of this assumptions checklist, after the proof of Lemma 8.2.15. Assumptions (5.8)–(5.11) are obtained in the very same way as we did in the proof of Theorem 8.2.1.

Checking (5.12) requires some more work. Let m be in $M_{\mathcal{I}}$. Consider a curve $m(s)$ on $M_{\mathcal{I}}$, such that $m(0) = m$. The curve $r_{\mathcal{I},t}\big(m(s)\big)$ lies on $\mathcal{D}_{B_t}$. In Lemma 8.2.14 we proved that $V_{\mathcal{I}}^{\perp}$ can be identified with $T_{r_{\mathcal{I},t}m(s)}\Lambda_{I(B_t)} \ominus T_{r_{\mathcal{I},t}(m(s))}\mathcal{D}_{B_t}$ for all small s. Hence, for any w in the unit sphere of $V_{\mathcal{I}}^{\perp}$ and λ positive, the curve $\exp_{r_{\mathcal{I},t}(m(s))}(\lambda w)$ on $\Lambda_{I(B_t)}$ is well defined. Its tangent vector field at $s=0$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}s}\exp_{r_{\mathcal{I},t}(m(s))}(\lambda w)\Big|_{s=0}$$

$$= \frac{\mathrm{d}}{\mathrm{d}s}\Bigl(\cos\Bigl(\frac{\lambda|w|}{\rho_t}\Bigr) r_{\mathcal{I},t}\bigl(m(s)\bigr) + \sin\Bigl(\frac{\lambda|w|}{\rho_t}\Bigr)\rho_t \frac{w}{|w|}\Bigr)\Big|_{s=0}$$

$$= \cos\Bigl(\frac{\lambda|w|}{\rho_t}\Bigr) r_{\mathcal{I},t,*}(m)m'(0) \,. \qquad (8.2.13)$$

Let $\tilde p = \exp_{r_{\mathcal{I},t}(m(s))}(\lambda w)$. Since $\pi_{B_t}(\tilde p) = r_{\mathcal{I},t}\bigl(m(s)\bigr)$ provided $\tilde p$ stays in $\omega_{B_t,r_{\mathcal{I},t}(m)}$, and since the orthocomplement of $\ker \pi_{B_t,*}(\tilde p)$ has dimension k, this orthocomplement can be identified as

$$\bigl(\ker \pi_{B_t,*}(\tilde p)\bigr)^{\perp} \equiv r_{\mathcal{I},t,*}(m)T_m M_{\mathcal{I}} = T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t} \,.$$

Moreover, (8.2.13) implies that for u in $T_{r_{\mathcal{I},t}(m)}\mathcal{D}_{B_t}$,

$$\pi_{B_t,*}(p)u = u\Big/ \cos\Bigl(\frac{\lambda|w|}{\rho}\Bigr) \,.$$

In other words, the restriction of $\pi_{B_t,*}(p)$ to $\bigl(\ker \pi_{B_t,*}(p)\bigr)^{\perp}$ is the map

$$\mathrm{Id}_{\mathbb{R}^k}\Big/ \cos\Bigl(\frac{\lambda|w|}{\rho}\Bigr) \,.$$

Consequently, $J\pi_{B_t}(p) = \cos\bigl(\lambda|w|/\rho\bigr)^{-k} = 1 + o(1)$ uniformly in the range $\lambda \leqslant O(\log\log t) = o(\rho_t)$ and m in $\mathcal{D}_{B_t}$. This proves (5.12).

Before checking (5.13), we need to evaluate the candidate for the limiting integral, namely

$$\int_{\mathcal{D}_{B_t}} \frac{e^{-\tau_{B_t}}\mathrm{I}_{[0,c(t)]}(\tau_{B_t})}{|\mathrm{D}I|^{\frac{d-k+1}{2}}(\det G_{B_t})^{1/2}} \mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}}$$

— notice that we use the Riemannian measure on $\mathcal{D}_{B_t}$ and not that on $\mathcal{D}'_{B_t}$. Since $r_{\mathcal{I},t}(m)$ is in the image of $q_{\mathcal{I},t}(m)$ by the map $u \in \mathbb{R}^d \setminus \{\,0\,\} \mapsto \rho_t u/|u|$ whose differential at $q_{\mathcal{I},t}(m)$ is $\bigl(\rho_t/|q_{\mathcal{I},t}(m)|\bigr)\mathrm{Proj}_{q_{\mathcal{I},t}(m)^\perp}$, we have

$$\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}}\bigl(r_{\mathcal{I},t}(m)\bigr) = \mathrm{d}\mathcal{M}_{\mathcal{D}'_{B_t}}\bigl(q_{\mathcal{I},t}(m)\bigr)\bigl(1+o(1)\bigr) \,, \qquad (8.2.14)$$

in the range $|q_{\mathcal{I},t}(m)| \sim \rho_t$. This range includes that for which $\tau_{B_t}\bigl(r_{\mathcal{I},t}(m)\bigr)$ is less than $c(t)$, thanks to Lemmas 8.2.12 and 8.2.13. Furthermore,

$$\begin{aligned}\tau_{B_t}\bigl(r_{\mathcal{I},t}(m)\bigr) &= \tau_{B_t}\bigl(q_{\mathcal{I},t}(m)\bigr) + \frac{1}{2}\Bigl(\bigl|q_{\mathcal{I},t}(m)\bigr|^2 - \rho_t^2\Bigr) \\ &= I\bigl(q_{\mathcal{I},t}(m)\bigr) - I(B_t) \\ &= \alpha\sum_{i\in\mathcal{I}} \log|m_i| - \alpha\log\gamma + o(1) \qquad \text{as } t\to\infty\end{aligned}$$

uniformly over $\tau_{B_t}\big(r_{t,\mathcal{I}}(m)\big) \leqslant c(t)$ — the first equality comes from (8.2.12) with $w = 0$; the second from the fact that $q_{\mathcal{I},t}(m)$ belongs to ∂B_t, and so $\tau_{B_t}\big(q_{\mathcal{I},t}(m)\big) = 0$; the third from (8.2.9) and Lemma 8.2.13.

Uniformly in $\tau_{B_t}\big(r_{\mathcal{I},t}(m)\big) \leqslant c(t)$, we have

$$\big|\mathrm{D}I\big(r_{\mathcal{I},t}(m)\big) \sim \big|\mathrm{D}I\big(q_{\mathcal{I},t}(m)\big)\big| \sim \sqrt{N\alpha\log t}\,.$$

Consequently,

$$\begin{aligned}
&\int_{\mathcal{D}_{B_t}} \frac{e^{-\tau_{B_t}}\mathrm{I}_{[0,c(t)]}(\tau_{B_t})}{|\mathrm{D}I|^{(d-k+1)/2}(\det G_{B_t})^{1/2}}\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}} \\
&\sim \int_{\mathcal{D}'_{B_t}} \frac{\gamma^\alpha\,\mathrm{I}_{[0,c(t)]}(\tau_{B_t})}{\prod_{i\in\mathcal{I}}|m_i|^\alpha (N\alpha\log t)^{(d-(N-1)+1)/4}(N\alpha\log t)^{-(d-N)/4}}\mathrm{d}\mathcal{M}_{D'_{B_t}} \\
&\sim \sum_{\mathcal{I}\in J(C)} \int_{M_\mathcal{I}} \frac{\gamma^\alpha}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\det\big(p_*(m)^{\mathrm{T}}p_*(m)\big)^{1/2}\mathrm{I}_{[0,c(t)]}\Big(\tau_{B_t}\big(r_{t,\mathcal{I}}(m)\big)\Big) \\
&\qquad\qquad \mathrm{d}\mathcal{M}_{M_\mathcal{I}}(m)\frac{1}{(N\alpha\log t)^{1/2}}\frac{1}{(\log t)^{(N-1)/2}} \qquad \text{as } t\to\infty\,.
\end{aligned}$$

Thus, the formula in Theorem 5.1 becomes

$$\begin{aligned}
&e^{-I(B_t)}(2\pi)^{(d-k-1)/2}\int_{\mathcal{D}_{B_t}}\frac{e^{-\tau_{B_t}}}{|\mathrm{D}I|^{(d-k+1)/2}(\det G_{B_t})^{1/2}}\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}} \\
&\sim \frac{(K_{s,\alpha}\alpha^{\alpha/2})^N}{\sqrt{N\alpha}}\frac{1}{t^{N\alpha/2}}\sum_{\mathcal{I}\in J(C)}\int_{M_\mathcal{I}}\frac{\det\big(p_*(m)^{\mathrm{T}}p_*(m)\big)^{1/2}}{\prod_{i\in\mathcal{I}}|m_i|^\alpha} \\
&\qquad\qquad \mathrm{I}_{[0,c(t)]}\Big(\tau_{B_t}\big(r_{t,\mathcal{I}}(m)\big)\Big)\mathrm{d}\mathcal{M}_{M_\mathcal{I}}(m)\,.
\end{aligned}$$

This formula is valid whenever α is positive. To remove the term $\mathrm{I}_{[0,c(t)]}\Big(\tau_{B_t}\big(r_{\mathcal{I},t}(m)\big)\Big)$ from the formula, it is enough to prove that for all $\mathcal{I}$ in $J(C)$,

$$\int_{M_\mathcal{I}}\frac{\det\big(p_*(m)^{\mathrm{T}}p_*(m)\big)^{1/2}}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\mathrm{d}\mathcal{M}_{M_\mathcal{I}}(m) < \infty\,, \tag{8.2.15}$$

because $\lim_{t\to\infty}c(t) = \infty$. This is where we need the assumption $\alpha > N/2$. For this purpose, we first derive a bound for $\det\big(p_*(m)^{\mathrm{T}}p_*(m)\big)$. The following result will be useful. Since $M_\mathcal{I}$ is in the complement of a neighborhood of the origin, it extends Lemma 8.2.11.

8.2.16. LEMMA. *There exists a constant K depending only on the matrix C such that for any $\mathcal{I}$ in $J(C)$, any m in $M_{\mathcal{I}}$ and i in $\mathcal{I}$,*

$$\frac{|m|}{K} \leqslant |m_i| \leqslant K|m|\,.$$

Proof. Assume that the lower bound were false, and that, for instance, $\inf_{m\in M_{\mathcal{I}}} |m_i|/|m| = 0$ for some i in $\mathcal{I}$ and $\mathcal{I}$ in $J(C)$. Then, there exists a sequence $m(n)$ in $M_{\mathcal{I}}$ such that $\lim_{n\to\infty} |m_i(n)|/|m(n)| = 0$. Given Lemma 8.2.11, this forces $\lim_{n\to\infty} |m(n)| = \infty$. Set $m^i(n) = m(n) - m(n)e_i = \mathrm{Proj}_{e_i^\perp} m(n)$. Define $s(n) = |m^i(n)|$ and $v(n) = m^i(n)/|m^i(n)|$. Since $|m(n)|$ tends to infinity with n, the condition $m_i(n)/|m(n)| = o(1)$, ensures that $|m^i(n)|$ does not vanish for n large enough. Thus $v(n)$ is well defined for n large enough. Moreover, $|m_i(n)| = o\bigl(s(n)\bigr)$ as n tends to infinity. We then have

$$\begin{aligned}
1 &= \langle Cm(n)\,, m(n)\rangle \\
&= m_i(n)^2 C_{i,i} + m_i(n)s(n)\bigl\langle (C+C^{\mathrm{T}})e_i\,, v(n)\bigr\rangle + s(n)^2\langle Cv(n)\,, v(n)\rangle \\
&= o\bigl(s(n)\bigr)^2 + s(n)^2\langle Cv(n)\,, v(n)\rangle\,.
\end{aligned}$$

Since $v(n)$ is in the compact sphere $S_{N-1} \subset V_{\mathcal{I}\setminus\{\,i\,\}}$, we can assume, up to extracting a subsequence, that $v = \lim_{n\to\infty} v(n)$ exists. Then v belongs to $V_{\mathcal{I}\setminus\{\,i\,\}}$ and we must have $\langle Cv, v\rangle \sim s(n)^{-2} = o(1)$, i.e., $\langle Cv, v\rangle = 0$. This contradicts assumption (8.2.8).

The upper bound is trivial since $|m_i| \leqslant |m|$ anyway. ∎

8.2.17. LEMMA. *If $\mathcal{I}$ is in $J(C)$ and m belongs to $M_{\mathcal{I}}$, then*

$$0 \leqslant \det\bigl(p_*(m)^{\mathrm{T}} p_*(m)\bigr) \leqslant \alpha^{N-1}\Bigl(\frac{K}{|m|}\Bigr)^{2(N-1)}\,.$$

Proof. From Lemma 8.2.16, we infer

$$\sqrt{\alpha}\sum_{i\in\mathcal{I}} \frac{e_i\otimes e_i}{|m_i|} \leqslant \frac{\sqrt{\alpha}K}{|m|}\mathrm{Id}_{V_{\mathcal{I}}}$$

in the sense that the difference between the right and left hand sides is a nonnegative matrix. Consequently, since restriction preserves matrix ordering,

$$p_*(m)^{\mathrm{T}} p_*(m) \leqslant \frac{K^2\alpha}{|m|^2}\mathrm{Id}_{\{\,(C+C^{\mathrm{T}})m\,\}^\perp\cap V_{\mathcal{I}}}\,.$$

Taking the determinant preserves the ordering as well, and the result follows. ∎

We can now prove (8.2.15).

8.2.18. LEMMA. *If $\alpha > 2/N$, then*

$$\int_{M_{\mathcal{I}}} \frac{\det\bigl(p_*(m)^{\mathrm{T}} p_*(m)\bigr)^{1/2}}{\prod_{i\in\mathcal{I}} |m_i|^\alpha} \mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m) < \infty\,.$$

Proof. Using Lemmas 8.2.16 and 8.2.17, it is enough to prove that

$$\int_{M_{\mathcal{I}}} \frac{\mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m)}{|m|^{N-1+N\alpha}} < \infty\,.$$

First, using a change of variable in polar coordinate, we see that for any ϵ positive and $s > N$,

$$\int_{\substack{|x|\geqslant\epsilon\\ x\in\mathbb{R}^N}} \frac{\mathrm{d}x}{|x|^s} < \infty\,.$$

Take $\epsilon < \epsilon_0$, where ϵ_0 is as in Lemma 8.2.11. In $V_{\mathcal{I}}$ identified with $\mathbb{R}^N$, make the change of variable $x = \theta m$, with θ positive and m in $M_{\mathcal{I}}$. We claim that

$$\int_{\substack{|x|\geqslant\epsilon\\ x\in V_{\mathcal{I}}}} \frac{\mathrm{d}x}{|x|^s} \geqslant \int_{\theta\geqslant 1}\int_{m\in M_{\mathcal{I}}} \frac{\bigl|\mathrm{Proj}_{(T_m M_{\mathcal{I}})^\perp}(m)\bigr|}{\theta^s |m|^s} \mathrm{d}\mathcal{M}(m)\theta^{N-1}\mathrm{d}\theta\,.$$

To obtain the right hand side and prove the claim, we argue as follows. Let $m(t_1,\ldots,t_{N-1})$ be a parameterization of $M_{\mathcal{I}}$. The Jacobian of the transformation $x = \theta m(t_1,\ldots,t_{N-1})$ is

$$\det\Bigl(\frac{\partial x}{\partial t_1},\ldots,\frac{\partial x}{\partial t_{N-1}},\frac{\partial x}{\partial\theta}\Bigr) = \theta^{N-1}\det\Bigl(\frac{\partial m}{\partial t_1},\ldots,\frac{\partial m}{\partial t_{N-1}},m\Bigr)\,.$$

Using the multilinearity of the determinant, writing m as a component on $\mathrm{span}\Bigl\{\frac{\partial m}{\partial t_1},\ldots,\frac{\partial m}{\partial t_{N-1}}\Bigr\} = T_m M_{\mathcal{I}}$ and a component on $(T_m M_{\mathcal{I}})^\perp$, we obtain

$$\det\Bigl(\frac{\partial m}{\partial t_1},\ldots,\frac{\partial m}{\partial t_{N-1}},m\Bigr) = \det\Bigl(\frac{\partial m}{\partial t_1},\ldots,\frac{\partial m}{\partial t_{N-1}},\mathrm{Proj}_{(T_m M_{\mathcal{I}})^\perp} m\Bigr)\,.$$

Up to a sign, this last determinant is

$$|\mathrm{Proj}_{(T_mM_{\mathcal{I}})^\perp}m|\Bigl(\det\Bigl(\frac{\partial m}{\partial t_1},\ldots,\frac{\partial m}{\partial t_{N-1}}\Bigr)^{\mathrm{T}}\Bigl(\frac{\partial m}{\partial t_1},\ldots,\frac{\partial m}{\partial t_{N-1}}\Bigr)\Bigr)^{1/2}.$$

This proves our claim. If $s>N$, we obtain, after performing the integration in θ,

$$\infty > \int_{m\in M_{\mathcal{I}}}\frac{\bigl|\mathrm{Proj}_{(T_mM_{\mathcal{I}})^\perp}(m)\bigr|}{|m|^s}\mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m)\,.$$

Recall that $(T_mM_{\mathcal{I}})^\perp\cap V_{\mathcal{I}} = \mathrm{Proj}_{V_{\mathcal{I}}}(C+C^{\mathrm{T}})m\mathbb{R}$. Since $M_{\mathcal{I}}$ is included in $V_{\mathcal{I}}$, it follows that for any m belonging to $M_{\mathcal{I}}$,

$$\begin{aligned}\mathrm{Proj}_{(T_mM_{\mathcal{I}})^\perp}(m) &= \frac{\langle m,(C+C^{\mathrm{T}})m\rangle}{|\mathrm{Proj}_{V_{\mathcal{I}}}(C+C^{\mathrm{T}})m|^2}\,\mathrm{Proj}_{V_{\mathcal{I}}}(C+C^{\mathrm{T}})m\\ &= \frac{2}{|\mathrm{Proj}_{V_{\mathcal{I}}}(C+C^{\mathrm{T}})m|^2}\,\mathrm{Proj}_{V_{\mathcal{I}}}(C+C^{\mathrm{T}})m\,.\end{aligned}$$

Thus,

$$|\mathrm{Proj}_{(T_mM_{\mathcal{I}})^\perp}(m)| \geqslant \frac{2}{\|\mathrm{Proj}_{V_{\mathcal{I}}}(C+C^{\mathrm{T}})\|\,|m|}\geqslant\frac{1}{\|C\|\,|m|}\,.$$

Consequently, for $s>N$, the integral $\displaystyle\int_{m\in M_{\mathcal{I}}}\frac{\mathrm{d}\mathcal{M}(m)}{|m|^{s+1}}$ is finite. This proves Lemma 8.2.18. ∎

Putting all the pieces together, we are left with only assumption (5.13) to check. But this is plain from the calculation we did in Lemmas 8.2.16 and 8.2.18. This concludes the proof of Theorem 8.2.10. ∎

We can now state an analogue of Theorem 8.2.9, that is a result on the limiting behavior of X given $\langle CX,X\rangle\geqslant t$ as t tends to infinity. For $\mathcal{I}$ in $J(C)$, Lemma 8.2.11 implies that the vector $\bigl(\mathrm{sign}(m_i)\bigr)_{i\in\mathcal{I}}$ is constant on the connected components of $M_{\mathcal{I}}$. Hence, to a connected component $\mathcal{C}$ of $M_{\mathcal{I}}$, we can associate a unique unit vector

$$\epsilon_{\mathcal{C}} = N^{-1/2}\sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)e_i\,,$$

where m is any point in $\mathcal{C}$. Moreover, for $\alpha > 2/N$ and $\tilde{G}(m)$ as defined in Theorem 8.2.10, the number

$$\mu_{\mathcal{C}} = \int_{\mathcal{C}} \frac{\big(\det \tilde{G}(m)\big)^{1/2}}{\prod_{i\in\mathcal{I}} |m_i|^\alpha} \mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m)$$

is finite.

8.2.19. THEOREM. *Under the assumptions of Theorem 8.2.10, the distribution of the random vector* $\big(\mathrm{sign}(X_i)\log|X_i|\big)_{1\leqslant i\leqslant d}\big/\log\sqrt{t}$ *given* $\langle CX, X\rangle \geqslant t$ *converges weakly* to*

$$\nu = \sum_{\mathcal{C}} \mu_{\mathcal{C}}\delta_{\epsilon_{\mathcal{C}}} \Big/ \sum_{\mathcal{C}} \mu_{\mathcal{C}}\,,$$

where the sums are over all connected components $\mathcal{C}$ *of* $\bigcup_{\mathcal{I}\in J(C)} M_{\mathcal{I}}$.

Proof. We proceed in a similar way as we did for proving Theorem 8.2.9. Set $\lambda_{B_t} = \rho_t$. The numerator of the measure in (5.18) when looking for the conditional distribution of Y/ρ_t given $Y \in B_t$ is

$$\begin{aligned}\frac{\rho_t^{N-1}\exp\big(-\tau_{B_t}(\rho_t r)\big)}{\rho_t^{(d-(N-1)+1)/2}(N\alpha\log t)^{-(d-N)/4}}\,&\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}/\rho_t}(r)\\ &\sim (N\alpha\log t)^{N/2}\exp\big(-\tau_{B_t}(\rho_t r)\big)\,\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}/\rho_t}(r)\,.\end{aligned}$$

Let f be a positive continuous function. Then (8.2.14) implies

$$\begin{aligned}\int f(r)\exp\big(-\tau_{B_t}(\rho_t r)\big)\,&\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}/\rho_t}(r)\\ &\sim \int f(r)\exp\big(-\tau_{B_t}(\rho_t r)\big)\,\mathrm{d}\mathcal{M}_{\mathcal{D}'_{B_t}/\rho_t}(r)\end{aligned}$$

as t tends to infinity. Using the parameterization of $\mathcal{D}_{B_t}$ and the expression for $\tau_{B_t(r_{\mathcal{I},t}(m))}$ obtained from equation (8.2.14), this last integral is also asymptotically equivalent to

$$\begin{aligned}\sum_{\mathcal{I}\in J(C)}\sum_{\mathcal{C}}\int_{\mathcal{C}} f\Big(\frac{r_{\mathcal{I},t}(m)}{\rho_t}\Big)\frac{\gamma^\alpha}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\times&\\ \det\big(p_*(m)^{\mathrm{T}}p_*(m)\big)^{1/2}\mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m)\,,&\qquad(8.2.16)\end{aligned}$$

where the second summation is over all connected components $\mathcal{C}$ of $M_{\mathcal{I}}$. But if m is in a connected component $\mathcal{C}$, then

$$\lim_{t\to\infty}\frac{r_{\mathcal{I},t}(m)}{\rho_t} = \epsilon_{\mathcal{C}}\,,$$

uniformly in the range of m's such that $q_{\mathcal{I},t}(m)$ belongs to $\mathcal{D}'_{B_t}$ — i.e., $\max_{i\in\mathcal{I}}\log|m_i|$ is less than $(\log t)^{1/4}$. Consequently, as t tends to infinity, (8.2.16) converges to

$$\sum_{\mathcal{I}\in J(C)}\sum_{\mathcal{C}\in M_{\mathcal{I}}} f(\epsilon_{\mathcal{C}})\gamma^\alpha\mu_{\mathcal{C}} = \gamma^\alpha\sum_{\mathcal{C}}\mu_{\mathcal{C}}\int f\,\mathrm{d}\nu\,.$$

Since the measure involved in (5.18) is normalized, we proved that the probability measure with density proportional to

$$\frac{\gamma^k\exp\big(-\tau_{B_t}(\rho_t q)\big)}{|\mathrm{D}I(\rho_t q)|^{\frac{d-k+1}{2}}\big(\det G_{B_t}(\rho_t q)\big)^{\frac12}}\,\mathrm{d}\mathcal{M}_{\mathcal{D}_{B_t}/\rho_t}(q)$$

converges weakly* to ν. Thus, (5.18) holds.

Assumption (5.19) holds trivially. Assumption (5.20) follows from the fact that ρ_t is of order $\sqrt{\log t}$, while for p in $\mathcal{D}_{B_t}$, the set $\pi_{B_t}^{-1}(p)\cap\underline{B}_{t,M}$ is in an $O(\log\log t)$-neighborhood of $\mathcal{D}_{B_t}$ thanks to Lemma 8.2.13.

Applying Corollary 5.3, we deduce that the conditional distribution of $Y/\sqrt{N\alpha\log t}$ given $Y\in B_t$ converges weakly* to ν.

Using the Skorokhod representation Theorem as we did in proving Theorem 8.2.9, Theorem 8.2.19 follows from

$$\begin{aligned} S_\alpha^{\leftarrow}\circ\Phi\Big(\epsilon_{\mathcal{C}}\sqrt{N\alpha\log t}\big(1+o(1)\big)\Big) \\ &= \sqrt{N}\epsilon_{\mathcal{C}}S_\alpha^{\leftarrow}\circ\Phi\Big(\sqrt{\alpha\log t}\big(1+o(1)\big)\Big) \\ &= \sqrt{N}\epsilon_{\mathcal{C}}\exp\Big(\log\sqrt{t}\big(1+o(1)\big)\Big) \end{aligned}$$

as t tends to infinity, thanks to A.1.6. ∎

One way to interpret Theorem 8.2.19 is to say that X given $\langle CX,X\rangle\geqslant t$ is distributed as $t^{(1/2)+o(1)}S$ where S is distributed according to ν. In other words, we obtained something like the first term of an expansion of X given $\langle CX,X\rangle\geqslant t$. Given the work done, a little extra effort will give us a second order term. That is, given $\langle CX,X\rangle\geqslant t$, and given that $\big(\mathrm{sign}(X_i)\log|X_i|\big)_{1\leqslant i\leqslant d}/\log\sqrt{t}$ is close to some $\epsilon_{\mathcal{C}}$, we can derive the limiting distribution of X. Refining the asymptotic analysis, we could obtain an asymptotic expansion of the distribution of X given $\langle CX,X\rangle\geqslant t$, as t tends to infinity, in term of successive conditional distributions. Such type of result is easier to phrase in term of random variables than in term of distributions.

8.2.20. THEOREM. *Under the assumptions of Theorem 8.2.10, the random variable X given $\langle CX, X\rangle \geqslant t$ can be represented as $\sqrt{t}m_g H\bigl(1+o(1)\bigr) + T_g + o(1)$, where the random variables g, m_g, H and t_g are as follows.*

The discrete random variable g has distribution ν.

Let $\mathcal{C}$ be a connected component of some $M_{\mathcal{I}}$. The conditional density of m_g given $g = \epsilon_{\mathcal{C}}$ is proportional to

$$\frac{|\mathrm{Proj}_{(T_{p(m)}p(\mathcal{C}))^\perp}\epsilon_{\mathcal{C}}|}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\det\tilde{G}(m)\,\mathrm{d}\mathcal{M}_{\mathcal{C}}(m)\,.$$

The random variable H has a Pareto distribution

$$P\{H\geqslant\lambda\} = \lambda^{-\alpha N}\,,\qquad \lambda\geqslant 1\,,$$

and is independent of g and m_g. Finally, T_g given g is a random vector in $V_{\mathcal{I}}^\perp$, with independent components, all having the original Student-like S_α distribution.

One way to read Theorem 8.2.20 is in terms of simulating X from its conditional distribution given $\langle CX, X\rangle \geqslant t$ for large t. We pick a connected component $\mathcal{C}$ with probability proportional to $\mu_{\mathcal{C}}$. Once $\mathcal{C}$ is picked, it lies in a unique subspace $V_{\mathcal{I}}$. In $V_{\mathcal{I}}$, we simulate m with distribution proportional to

$$\frac{|\mathrm{Proj}_{(T_{p(m)}p(\mathcal{C}))^\perp}\epsilon_{\mathcal{C}}|}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\det\tilde{G}(m)\,\mathrm{d}\mathcal{M}_{\mathcal{C}}(m)\,.$$

Next, we simulate H with the Pareto distribution. Then the part of X in $V_{\mathcal{I}}$ is $\sqrt{t}mH\bigl(1+o(1)\bigr)$ as t tends to infinity. The part of X in $V_{\mathcal{I}}^\perp$ is a random vector with independent and identically distributed components from the initial Student-like distribution, up to an additive term of order $o(1)$ as t tends to infinity. We will explicitly calculate the norm of $\mathrm{Proj}_{T_{p(m)}}p(\mathcal{C})\epsilon_{\mathcal{C}}$ at the end of the proof of Theorem 8.2.20.

Proof of Theorem 8.2.20. The intuition behind the proof is extremely simple given all that we did. Looking at $Y = \Phi^{\leftarrow}\circ S_\alpha(X)$, we want to obtain an approximation of Y given $Y\in B_t$, and invert it to obtain one of $X = S_\alpha^{\leftarrow}\circ\Phi(Y)$ given $\langle CX, X\rangle \geqslant t$. Lemma 8.2.12 asserts that

the points in B_t near $Q(t)\sqrt{N}\epsilon_{\mathcal{C}}$ are of the form $\psi\big(q_{\mathcal{I},t}(m,v),s\big)$. Set $w=\Phi^{\leftarrow}\circ S_\alpha(v)$, and define

$$\tilde q_{\mathcal{I},t}(m,w)=\sum_{i\in\mathcal{I}}\mathrm{sign}(m_i)\Big(Q(t)+\frac{\sqrt{\alpha}\log|m_i|}{\sqrt{\log t}}\Big)e_i+w\,,\qquad w\in V_{\mathcal{I}}^{\perp}\,.$$

Formula (8.2.9), the definition of the normal flow, and Lemma 8.2.13 show that the normal density at $\psi\big(\tilde q_{\mathcal{I},t}(m,w),s\big)$ is

$$\begin{aligned}\exp\Big[-I\Big(\psi_s\big(\tilde q_{\mathcal{I},t}(m,w)\big)\Big)\Big]&=\exp\Big[-I\big(\tilde q_{\mathcal{I},t}(m,w)\big)-s\Big]\\&=\frac{e^{-R(t)}e^{-s}e^{-|w|^2/2}}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\big(1+o(1)\big)\end{aligned}\qquad(8.2.17)$$

In this expression, the term e^{-s} is an exponential density, the term $e^{-|w|^2/2}$ is a Gaussian one, and the term $\prod_{i\in\mathcal{I}}|m_i|^\alpha$ will give us another density. We can parameterize points y of B_t in term of s,w,m. Then, we interpret these parameters as random variables. This will give a representation of the random variable Y given $Y\in B_t$, and we will pull back this representation to X.

However, we need to be careful with the different scales. For m in a connected component $\mathcal{C}$ of $M_{\mathcal{I}}$,

$$\sqrt{\log t}\,\mathrm{Proj}_{V_{\mathcal{I}}^{\perp}}\big(\tilde q_{\mathcal{I},t}(m,w)-Q(t)\sqrt{N}\epsilon_{\mathcal{C}}\big)=p(m)\,,$$

while

$$\mathrm{Proj}_{V_{\mathcal{I}}}\tilde q_{\mathcal{I},t}(m,w)=w\,.$$

Thus, the part of y in $V_{\mathcal{I}}$ should be centered and rescaled by $\sqrt{\log t}$, while that in $V_{\mathcal{I}}^{\perp}$ is already of order 1.

To proceed rigorously, let f be a nonnegative smooth function defined on $\mathbb{R}^N\times\mathbb{R}^{d-N}$, with compact support.

For some $\mathcal{I}$ in $J(C)$ and a connected component $\mathcal{C}$ of $M_{\mathcal{I}}$, let us evaluate the integral

$$\int_{B_t}f\Big(\sqrt{\log t}\,\mathrm{Proj}_{V_{\mathcal{I}}}\big(y-Q(t)\sqrt{N}\epsilon_{\mathcal{C}}\big),\mathrm{Proj}_{V_{\mathcal{I}}^{\perp}}y\Big)e^{-I(y)}\,\mathrm{d}y\,.\qquad(8.2.18)$$

Divided by $P\{\,Y\in B_t\,\}$, this integral will give us the limiting conditional behavior of Y given $Y\in B_t$, after proper normalization.

If the projection of y onto $V_{\mathcal{I}}$ is not in a neighborhood of $Q(t)\sqrt{N}\epsilon_{\mathcal{C}}$, then $\sqrt{\log t}\,\mathrm{Proj}_{V_{\mathcal{I}}}\big(y-Q(t)\sqrt{N}\epsilon_{\mathcal{C}}\big)$ diverges as t tends to infinity. Since f is compactly supported, such points y do not contribute to

the integral for large t. From the preceding, since f is bounded, we can also restrict the range of integration for those y's such that $I(y) \leqslant I(B_t) + c(t)$. Let us make the change of variable

$$y = \psi\big(\tilde{q}_{\mathcal{I},t}(m,w), s\big) = \sqrt{1 + \frac{2s}{|\tilde{q}_{\mathcal{I},t}(m,w)|^2}}\, \tilde{q}_{\mathcal{I},t}(m,w)\,.$$

Notice that

$$|\tilde{q}_{\mathcal{I},t}(m,w)|^2 \geqslant |q_{\mathcal{I},t}(m)|^2 \sim N\alpha \log t$$

as t tends to infinity, and uniformly in the range $\tau_{B_t}\big(\tilde{q}_{\mathcal{I},t}(m)\big) \leqslant c(t)$. Therefore,

$$y = \tilde{q}_{\mathcal{I},t}(m,w) + \frac{s}{\sqrt{N\alpha \log t}}\epsilon_{\mathcal{C}} + o\Big(\frac{1}{\sqrt{\log t}}\Big)$$

as t tends to infinity, uniformly in the range $I(y) \leqslant I(B_t) + c(t)$. Consequently, in that range of y's,

$$\sqrt{\log t}\, \mathrm{Proj}_{V_{\mathcal{I}}}\big(y - Q(t)\sqrt{N}\epsilon_{\mathcal{C}}\big) = p(m) + \frac{s}{\sqrt{N\alpha}}\epsilon_{\mathcal{C}} + o(1) \qquad \text{as } t \to \infty\,.$$

Using (8.2.17), we can rewrite (8.2.18) as

$$\int_{s\geqslant 0}\int_{m\in M_{\mathcal{I}}}\int_{w\in V_{\mathcal{I}}^{\perp}} \mathrm{I}_{[0,I(B_t)+c(t)]}\big(I(y)\big) f\Big(p(m) + \frac{s}{\sqrt{N\alpha}}\epsilon_{\mathcal{C}}, w\Big) \times \\ \frac{e^{-R(t)}}{\prod_{i\in\mathcal{I}} |m_i|^{\alpha}} e^{-s} e^{-|w|^2/2} J(m,w,s) \mathrm{d}\mathcal{M}_{M_{\mathcal{I}}}(m)\, \mathrm{d}w\, \mathrm{d}s\, \big(1 + o(1)\big)\,,$$

where $J(m,w,s)$ is a Jacobian term. To calculate it, we first have

$$\frac{\partial y}{\partial s} = \frac{\tilde{q}_{\mathcal{I},t}(m,w)}{|\tilde{q}_{\mathcal{I},t}(m,w)|^2 \sqrt{1 + \frac{2s}{|\tilde{q}_{\mathcal{I},t}(m,w)|^2}}} = \frac{\epsilon_{\mathcal{C}}}{\sqrt{N\alpha\log t}}\big(1 + o(1)\big)$$

as t tends to infinity, uniformly in s in any compact set of $\mathbb{R}$ and m, w such that $I(y) \leqslant I(B_t) + c(t)$.

The explicit expression of $\psi_s(q)$ for the Gaussian distribution gives

$$\psi_{s,*}(q) = \sqrt{1 + \frac{2s}{|q|^2}}\,\mathrm{Id} - \frac{2s}{|q|^2}\frac{q\otimes q}{|q|^2}\frac{1}{\sqrt{1 + \frac{2s}{|q|^2}}}\,.$$

Let $m = m(u_1, \ldots, u_{N-1})$ be a local parameterization of $\mathcal{C}$. Then

$$\frac{\partial \tilde{q}_{\mathcal{I},t}}{\partial u_j}(m,w) = \frac{1}{\sqrt{\log t}}\, p_*(m) \frac{\partial m}{\partial u_j} \, .$$

Consequently

$$\frac{\partial y}{\partial u_j} = \psi_{s,*}\big(\tilde{q}_{\mathcal{I},t}(m,w)\big) \frac{\partial \tilde{q}_{\mathcal{I},t}(m,w)}{\partial u_j} = \frac{1}{\sqrt{\log t}} p_*(m) \frac{\partial m}{\partial u_j}\big(1 + o(1)\big) \, ,$$

uniformly in y such that $I(y) \leqslant I(B_t) + c(t)$. Finally, since $\dfrac{\partial}{\partial w_j}\tilde{q}_{\mathcal{I},t}(m,w) = e_j$, we have

$$\frac{\partial y}{\partial w_j} = \psi_{s,*}\big(\tilde{q}_{\mathcal{I},t}(m,w)\big) e_j = e_j + o(1) \qquad \text{as } t \to \infty \, .$$

Define the $(d \times d)$-matrix with columns indexed by i and rows indexed by j,

$$F = \begin{array}{c} \\ {\scriptstyle N-1} \\ {\scriptstyle 1} \\ {\scriptstyle d-N} \end{array} \!\!\! \begin{array}{c} \begin{array}{ccc} \quad {\scriptstyle N-1} \quad & \quad {\scriptstyle 1} \quad & \quad {\scriptstyle d-N} \quad \end{array} \\ \begin{pmatrix} \left\langle \frac{\partial y}{\partial u_i}, \frac{\partial y}{\partial u_j} \right\rangle_{i,j} & \left\langle \frac{\partial y}{\partial u_i}, \frac{\partial y}{\partial s} \right\rangle_i & \left\langle \frac{\partial y}{\partial u_i}, \frac{\partial y}{\partial w_j} \right\rangle_{i,j} \\ \left\langle \frac{\partial y}{\partial s}, \frac{\partial y}{\partial u_j} \right\rangle_j & \left| \frac{\partial y}{\partial s} \right|^2 & \left\langle \frac{\partial y}{\partial s}, \frac{\partial y}{\partial w_j} \right\rangle_j \\ \left\langle \frac{\partial y}{\partial w_i}, \frac{\partial y}{\partial u_j} \right\rangle_{i,j} & \left\langle \frac{\partial y}{\partial w_j}, \frac{\partial y}{\partial s} \right\rangle_i & \left\langle \frac{\partial y}{\partial w_i}, \frac{\partial y}{\partial w_j} \right\rangle_{i,j} \end{pmatrix} \end{array} .$$

This ensures that $J(m,w,s) = (\det F)^{1/2}$. Since $\partial y / \partial w_i$ is orthogonal to $V_{\mathcal{I}}$, while $\partial y / \partial u_i$ and $\partial y / \partial s$ are in $V_{\mathcal{I}}$, and since $\partial y / \partial w_j$ is roughly e_j, the determinant of F is equal to $o(\log t)^{-N}$ plus the determinant of the upper left $N \times N$ block of F, that is

$$\det \begin{pmatrix} \frac{1}{\log t} \left\langle p_*(m) \frac{\partial m}{\partial u_i}, p_*(m) \frac{\partial m}{\partial u_j} \right\rangle_{i,j} & \frac{1}{\sqrt{N\alpha} \log t} \left\langle p_*(m) \frac{\partial m}{\partial u_i}, \epsilon_{\mathcal{C}} \right\rangle_i \\ \frac{1}{\sqrt{N\alpha} \log t} \left\langle \epsilon_{\mathcal{C}}, p_*(m) \frac{\partial m}{\partial u_j} \right\rangle_j & \frac{1}{N\alpha \log t} \end{pmatrix}$$

— in this determinant, i, j run over $1, \ldots, N-1$. This determinant is equal to

$$\begin{aligned} &\frac{1}{N\alpha (\log t)^N} \det \begin{pmatrix} \left\langle p_*(m) \frac{\partial m}{\partial u_i}, p_*(m) \frac{\partial m}{\partial u_j} \right\rangle_{i,j} & \left\langle p_*(m) \frac{\partial m}{\partial u_i}, \epsilon_{\mathcal{C}} \right\rangle_i \\ \left\langle \epsilon_{\mathcal{C}}, p_*(m) \frac{\partial m}{\partial u_j} \right\rangle_j & 1 \end{pmatrix} \\ &\qquad = \frac{|\mathrm{Proj}_{(T_{p(m)} p(\mathcal{C}))^\perp} \epsilon_{\mathcal{C}}|^2}{N\alpha (\log t)^N} \det\left(\left\langle p_*(m) \frac{\partial m}{\partial u_i}, p_*(m) \frac{\partial m}{\partial u_j} \right\rangle_{i,j} \right) . \end{aligned}$$

Consequently, up to $o(\log t)^{-N}$,

$$\mathrm{d}y = \frac{|\mathrm{Proj}_{(T_{p(m)}p(\mathcal{C}))^\perp}\epsilon_{\mathcal{C}}|}{\sqrt{N\alpha}(\log t)^{N/2}} \det\bigl(p_*(m)^{\mathrm{T}} p_*(m)\bigr)^{1/2} \,\mathrm{d}\mathcal{M}_{\mathcal{C}}(m)\,\mathrm{d}s\,\mathrm{d}w\,.$$

Let us write

$$p(\mathcal{C}) = \{\, p(m) \,:\, m \in \mathcal{C}\,\}\,,$$

the image of a connected component $\mathcal{C}$ by $p(\cdot)$. If follows from our calculation that that the integral in (8.2.18) is

$$\int f\Bigl(p(m) + \frac{s}{\sqrt{N\alpha}}\epsilon_{\mathcal{C}}, w\Bigr) \frac{e^{-R(t)}}{\prod_{i\in\mathcal{I}}|m_i|^\alpha} e^{-s} e^{-|w|^2/2}\,\mathrm{d}s\,|\mathrm{Proj}_{(T_{p(m)}p(\mathcal{C}))^\perp}\epsilon_{\mathcal{C}}|$$
$$\det\bigl(p_*^{\mathrm{T}}(m)p_*(m)\bigr)^{1/2}\,\mathrm{d}\mathcal{M}_{\mathcal{C}}(m)\,\mathrm{d}w\,\frac{1}{\sqrt{N\alpha}(\log t)^{N/2}}\bigl(1+o(1)\bigr)\,.$$

Elementary algebra shows that

$$\frac{e^{-R(t)}}{\sqrt{N\alpha}(\log t)^{N/2}} = \frac{(K_{s,\alpha}\alpha^{\alpha/2})^N}{t^{N\alpha/2}\sqrt{N\alpha}} \frac{1}{(2\pi)^{(d-N)/2}}\,.$$

Consequently, as t tends to infinity, (8.2.18) is equivalent to

$$\int f\Bigl(p(m) + \frac{s}{\sqrt{N\alpha}}\epsilon_{\mathcal{C}}, w\Bigr) \prod_{i\in\mathcal{I}} |m_i|^{-\alpha} e^{-s} \frac{e^{-|w|^2/2}}{(2\pi)^{(d-N)/2}} \times$$
$$|\mathrm{Proj}_{(T_{p(m)}p(\mathcal{C}))^\perp}\epsilon_{\mathcal{C}}| \det\bigl(p_*(m)^{\mathrm{T}}p_*(m)\bigr)^{1/2} \mathrm{d}s\,\mathrm{d}\mathcal{M}_{\mathcal{C}}(m)\,\mathrm{d}w\,.$$

Combining this estimate with that for $P\{\, X \in B_t \,\}$ given by Theorem 8.2.10, the conditional distribution of

$$\Bigl(\sqrt{\log t}\,\mathrm{Proj}_{V_{\mathcal{I}}}\bigl(Y - Q(t)\sqrt{N}\epsilon_{\mathcal{C}}\bigr), \mathrm{Proj}_{V_{\mathcal{I}}^\perp}(Y)\Bigr) \qquad \text{given } Y \in B_t$$

and Y in a neighborhood of $\epsilon_{\mathcal{C}}$ converges weakly* to that of $\bigl(p(M) + (S\epsilon_{\mathcal{C}}/\sqrt{N\alpha}), W\bigr)$ where M, S and W are independent with respective densities proportional to

$$|\mathrm{Proj}_{T_{p(m)}(p(\mathcal{C}))^\perp}\epsilon_{\mathcal{C}}| \frac{\det\bigl(p_*(m)^{\mathrm{T}}p_*(m)\bigr)^{1/2}}{\prod_{i\in\mathcal{I}}|m_i|^\alpha}\,\mathrm{d}\mathcal{M}_{p(\mathcal{C})}(p)\,,$$
$$e^{-s}\mathrm{d}s\,, \qquad \text{and} \qquad \mathcal{N}(0, \mathrm{Id}_{V_{\mathcal{I}}^\perp})\,.$$

In particular, S has an exponential density.

Using Skorokhod's representation theorem and up to changing the versions of the random variables, given $Y \in B_t$ and Y in the neighborhood of $Q(t)\sqrt{N}\epsilon_{\mathcal{C}}$, we have

$$\begin{aligned} \mathrm{Proj}_{V_{\mathcal{I}}} Y &= Q(t)\sqrt{N}\epsilon_{\mathcal{C}} + \frac{1}{\sqrt{\log t}}\Big(p(M) + \frac{S}{\sqrt{N\alpha}}\epsilon_{\mathcal{C}}\Big) + o(\log t)^{-1/2} \\ \mathrm{Proj}_{V_{\mathcal{I}}^{\perp}} Y &= W + o(1)\,. \end{aligned}$$

Then, given $X \in A_t$, we have

$$X = S_\alpha^{\leftarrow} \circ \Phi(\mathrm{Proj}_{V_{\mathcal{I}}} Y) + S_\alpha^{\leftarrow} \circ \Phi\big(W + o(1)\big)\,.$$

The term $T_{\mathcal{I}} = S_\alpha^{\leftarrow} \circ \Phi\big(W + o(1)\big)$ is asymptotically a random vector in $V_{\mathcal{I}}^{\perp}$ with independent coefficients having a Student-like S_α distribution. Then, for i in $\mathcal{I}$ and t large enough, Lemma A.1.6 yields

$$\begin{aligned} &S_\alpha^{\leftarrow} \circ \Phi(\langle \mathrm{Proj}_{V_{\mathcal{I}}} Y, e_i\rangle) \\ &= S_\alpha^{\leftarrow} \circ \Phi\Big(Q(t)\sqrt{N}\langle \epsilon_{\mathcal{C}}, e_i\rangle + \frac{1}{\sqrt{\log t}}\Big\langle p(M) + \frac{S}{\sqrt{N\alpha}}\epsilon_{\mathcal{C}}, e_i\Big\rangle \\ &\qquad\qquad + o(\log t)^{-1/2}\Big) \\ &= \mathrm{sign}(m_i) S_\alpha^{\leftarrow} \circ \Phi\Big(Q(t) + \frac{\mathrm{sign}(m_i)}{\sqrt{\log t}}\langle p(M), e_i\rangle + \frac{S}{N\sqrt{\alpha}} + o(\log t)^{-1/2}\Big) \\ &= \sqrt{t}\,\mathrm{sign}(m_i)\exp\Big(\frac{S}{\alpha N} + \frac{\mathrm{sign}(m_i)}{\sqrt{\alpha}}\langle p(M), e_i\rangle + o(1)\Big)\,. \end{aligned}$$

If both M and m belong to $\mathcal{C}$, then

$$\alpha^{-1/2}\mathrm{sign}(m_i)\langle p(M), e_i\rangle = \log|\langle M, e_i\rangle|\,.$$

Therefore,

$$\begin{aligned} X &= \sqrt{t}\sum_{i\in\mathcal{I}} e^{S/(\alpha N)}\langle M, e_i\rangle e_i\big(1 + o(1)\big) + T_{\mathcal{I}} + o(1) \\ &= \sqrt{t}e^{S/(\alpha N)} M\big(1 + o(1)\big) + T_{\mathcal{I}} + o(1) \end{aligned}$$

To conclude the proof, notice first that $e^{S/(\alpha N)}$ has a Pareto distribution, since

$$P\Big\{\exp\Big(\frac{w}{\alpha N}\Big) \geqslant x\Big\} = P\{\, w \geqslant \alpha N \log x\,\} = e^{-\alpha N \log x} = x^{-\alpha N}\,.$$

Finally, as announced, let us calculate the norm of the projection of $\epsilon_{\mathcal{C}}$ onto the orthocomplement of the tangent space of $p(\mathcal{C})$ at $p(m)$. Since

$$p_*(m) = \sum_{i\in\mathcal{I}} \frac{e_i \otimes e_i}{|m_i|}\,,$$

the tangent space of $p(\mathcal{C})$ can be identified as

$$\begin{aligned} T_{p(m)}p(\mathcal{C}) &= \sqrt{\alpha}\sum_{i\in\mathcal{I}} \frac{e_i \otimes e_i}{|m_i|} T_m\mathcal{C} \\ &= \sqrt{\alpha}\sum_{i\in\mathcal{I}} \frac{e_i \otimes e_i}{|m_i|} \{(C+C^{\mathrm{T}})m\}^{\perp}\,. \end{aligned}$$

Defining $v(m) = \sum_{i\in\mathcal{I}} |m_i|\langle (C+C^{\mathrm{T}})m, e_i\rangle e_i$, we have

$$T_{p(m)}p(\mathcal{C}) = V_{\mathcal{I}} \cap \{v\}^{\perp}\,.$$

Consequently, using that $\langle Cm, m\rangle = 1$, we have

$$\begin{aligned} |\mathrm{Proj}_{T_{p(m)}p(\mathcal{C})^{\perp}}\epsilon_{\mathcal{C}}| &= \frac{\langle v(m), \epsilon_{\mathcal{C}}\rangle}{|v(m)|} \\ &= \frac{1}{\sqrt{N}} \frac{\sum_{i\in\mathcal{I}} mi\langle (C+C^{\mathrm{T}})m, e_i\rangle}{\big(\sum_{i\in\mathcal{I}} m_i^2\langle (C+C^{\mathrm{T}})m, e_i\rangle^2\big)^{1/2}} \\ &= \Big(N\sum_{i\in\mathcal{I}} m_i^2\langle (C+C^{\mathrm{T}})m, e_i\rangle^2\Big)^{1/2}\,. \end{aligned}$$

This concludes the proof of Theorem 8.2.20. ∎

Theorems 8.2.1 and 8.2.10 do not settle completely the tail behavior of $\langle CX, X\rangle$. For instance, it may happen that the largest diagonal term of C vanishes, or that $N(C) > 1$ but (8.2.8) does not hold. The situation not covered assumes that

$\langle Cu, u\rangle = 0$ for some nonzero u in $V_{\mathcal{I}}$ and some $\mathcal{I}$ in T, of cardinality strictly less than $N(C)$. (8.2.19)

The technique developed in this section may work in this situation; but we will see in the next section that, under (8.2.19), some extra complication is added. For the time being, we will only prove a rather weak result. It will be useful for the statistical applications developed in chapter 9.

Denote $N_0(C)$ the smallest cardinal of a set $\mathcal{I}$ such that (8.2.19) holds. Under (8.2.19), the next result asserts that the tail probability of $\langle CX, X\rangle$ is much lighter than $t^{-\alpha N_0(C)/2}$.

8.2.21. THEOREM. *Let X be a d-dimensional random vector with independent and identically distributed components having a Student like distribution with parameter α. Let C be a $d\times d$ matrix, and write $N_0 = N_0(C)$. Under (8.2.19), and if $\alpha > 2/N_0$,*

$$\lim_{t\to\infty} t^{\alpha N_0/2} P\{\,\langle CX, X\rangle \geqslant t\,\} = 0\,.$$

Proof. The statement is obvious if C is negative. Thus, we assume that C is not negative. For ϵ positive, denote $C_\epsilon = C + \epsilon\mathrm{Id}$. The proof relies on two very basic observations. The first one is that for any positive ϵ the matrix $C_\epsilon - C$ is positive. Consequently, for any t,

$$P\{\,\langle CX, X\rangle \geqslant t\,\} \leqslant P\{\,\langle C_\epsilon X, X\rangle \geqslant t\,\}\,.$$

Assumption (8.2.19) states that $N_0 < N(C)$. Let

$$J_0(C) = \{\,\mathcal{I}\in T\,:\,|\mathcal{I}| = N_0(C)\,,\,\exists u\in V_{\mathcal{I}}\setminus\{\,0\,\}\,,\,\langle Cu, u\rangle = 0\,\}\,.$$

The second observation is stated in the following lemma.

8.2.22. LEMMA. *If ϵ is positive and small enough, then $N(C_\epsilon) = N_0(C)$ and $J(C_\epsilon) = J_0(C)$. Moreover, C_ϵ satisfies assumption (8.2.8).*

Proof. Denote by T_{N_0} the set of all subsets of $\{\,1,\ldots,d\,\}$ of cardinal at most N_0. Assumption (8.2.19) is equivalent to the following. Whenever $\mathcal{I}$ is in $T_{N_0}\setminus J_0(C)$, the compression of C to $V_{\mathcal{I}}$ is negative; moreover, for any $\mathcal{I}$ in $J_0(C)$, this compression is nonpositive and there exists a nonzero $z_{\mathcal{I}}$ in $V_{\mathcal{I}}$ for which $\langle Cz_{\mathcal{I}}, z_{\mathcal{I}}\rangle = 0$. Consequently, (8.2.19) implies the existence of a positive ϵ_0 such that $\langle Cu, u\rangle < -\epsilon_0$ for any $\mathcal{I}$ in $T_{N_0}\setminus J_0(C)$ and any unit vector u of $V_{\mathcal{I}}$. If ϵ is positive, less than ϵ_0, and $\mathcal{I}$ is in $T_{N_0}\setminus J_0(C)$, the compression of C_ϵ to $V_{\mathcal{I}}$ is negative. Furthermore, if $\mathcal{I}$ is in $J_0(C)$, then $\langle C_\epsilon z_{\mathcal{I}}, z_{\mathcal{I}}\rangle = \epsilon|z_{\mathcal{I}}|^2$ is positive. This proves $N(C_\epsilon) = N_0$ and $J(C_\epsilon) = J_0(C)$ as well as C_ϵ satisfies (8.2.8). ∎

Let ϵ be positive and small enough so that the conclusions of Lemma 8.2.22 hold. For $\mathcal{I}$ in $J_0(C) = J(C_\epsilon)$, denote

$$M_{\epsilon,\mathcal{I}} = \{\,m\in V_{\mathcal{I}} : \langle C_\epsilon m, m\rangle = 1\,\}\,.$$

Using our two observations and Theorem 8.2.10, we infer that for any ϵ positive and small enough,

$$\begin{aligned}\limsup_{t\to\infty} & t^{\alpha N_0/2} P\{\, \langle CX, X\rangle \geqslant t\,\} \\ & \leqslant \frac{K_{s,\alpha}\alpha^{\alpha N_0/2}}{\sqrt{\alpha N_0}} \sum_{\mathcal{I}\in J_0(C)} \int_{M_{\epsilon,\mathcal{I}}} \frac{\det \tilde{G}(m)}{\prod_{i\in\mathcal{I}} |m_i|^\alpha} \mathrm{d}\mathcal{M}_{M_{\epsilon,\mathcal{I}}}(m)\,.\end{aligned}$$

To conclude the proof, we show that the above upper bound tends to 0 as ϵ tends to 0. For this purpose, we need a good description of $M_{\epsilon,\mathcal{I}}$. It will be helpful to have some understanding of the sets

$$Z_{\mathcal{I}}(C) = \{\, u \in V_{\mathcal{I}} \,:\, \langle Cu, u\rangle = 0\,\}\,, \qquad \mathcal{I} \in J_0(C)\,.$$

Despite the fact that these sets are defined by a quadratic equation, minimality of N_0 implies that they are linear spaces.

8.2.23. LEMMA. *Under (8.2.19), the set $Z_{\mathcal{I}}(C)$ is a one dimensional vector space. It coincides with the null eigensubspace in $V_{\mathcal{I}}$ of the compression of $C + C^{\mathrm{T}}$ to $V_{\mathcal{I}}$.*

Proof. Since the compression of C to $V_{\mathcal{I}}$ is nonpositive,

$$Z_{\mathcal{I}}(C) = \Bigl\{\, u \in V_{\mathcal{I}} \,:\, \langle Cu, u\rangle = \sup_{v\in V_{\mathcal{I}}} \langle Cv, v\rangle \,\Bigr\}\,.$$

The equality $2\langle Cu, u\rangle = \langle (C + C^{\mathrm{T}})u, u\rangle$, shows that $Z_{\mathcal{I}}(C)$ is the eigensubspace associated to the largest eigenvalue of the compression of $(C + C^{\mathrm{T}})$ to $V_{\mathcal{I}}$. Thus, $Z_{\mathcal{I}}(C)$ is indeed a linear space.

Let i be in $\mathcal{I} \in J_0(C)$ and u be a nonzero vector in $Z_{\mathcal{I}}(C)$. If $\langle u, e_i\rangle$ vanishes, u belongs to the (N_0-1)-dimensional subspace $V_{\mathcal{I}\setminus\{i\}}$, contradicting the minimality of $N(C)$. Consequently, no component of u in $V_{\mathcal{I}}$ vanishes.

Assume now that we can find two linearly independent vectors u and v in $Z_{\mathcal{I}}(C)$. Since none of their components in $V_{\mathcal{I}}$ vanish, there exists a linear combination of these two vector with at least one component in $V_{\mathcal{I}}$ vanishing. This linear combination is a vector which contradicts what we just showed. Consequently, there are not two linearly independent vectors in $Z_{\mathcal{I}}(C)$. Since $Z_{\mathcal{I}}(C)$ in nonempty by definition of N_0, this concludes the proof. ∎

It follows from Lemma 8.2.23 that the unit sphere of the space $Z_{\mathcal{I}}(C)$ is actually made of two points, z and $-z$.

To describe $M_{\epsilon,\mathcal{I}}$, we now introduce the sphere corresponding to the compression of $-C$ to the orthocomplement of $Z_{\mathcal{I}}(C)$ in $V_{\mathcal{I}}$, that is

$$S_{\mathcal{I}}(C) = \{\, v \in V_{\mathcal{I}} \,:\, \langle Cv, v\rangle = -1\,;\, v \perp Z_{\mathcal{I}}(C)\,\}\,.$$

Since $Z_{\mathcal{I}}(C)$ is the eigenspace associated to the simple null eigenvalue of $\mathrm{Proj}_{V_{\mathcal{I}}}(C + C^{\mathrm{T}})\big|_{V_{\mathcal{I}}}$, the set $S_{\mathcal{I}}(C)$ is a compact ellipsoid. We can now make $M_{\epsilon,\mathcal{I}}$ explicit.

8.2.24. LEMMA. *For any $\mathcal{I}$ in $J_0(C)$, for any ϵ positive and small enough,*

$$M_{\epsilon,\mathcal{I}} = \Big\{\, \frac{\eta}{\sqrt{\epsilon}}\sqrt{1 + s(1-\epsilon)}\, z + \sqrt{s}\, v \,:\, v \in S_{\mathcal{I}}(C)\,, \qquad s \geqslant 0\,,\, \eta \in \{\,-1, 1\,\}\Big\}\,,$$

where z is in the unit sphere of $Z_{\mathcal{I}}(C)$.

Proof. Let $\mathcal{I}$ be in $J_0(C)$. A point m in $V_{\mathcal{I}}$ belongs to $M_{\epsilon,\mathcal{I}}$ if and only if

$$1 = \langle Cm, m\rangle + \epsilon |m|^2\,. \tag{8.2.20}$$

The nonpositivity of C on $V_{\mathcal{I}}$ implies $\langle Cm, m\rangle \leqslant 0$. Therefore $|m|^2 \geqslant 1/\epsilon$. In particular, m is nonzero. Write $\lambda = |m|^2$ and $p = m/|m|$. Equality (8.2.20) and nonpositivity of C on $V_{\mathcal{I}}$ imply

$$0 \geqslant \langle Cp, p\rangle = -\epsilon + \frac{1}{\lambda} \geqslant -\epsilon\,. \tag{8.2.21}$$

The function $u \mapsto \langle Cu, u\rangle$ is continuous. Its restriction to the — compact — unit sphere of $V_{\mathcal{I}}$ is maximum exactly on $Z_{\mathcal{I}}(C)$. Consequently, if U is an arbitrary neighborhood of the unit sphere of $Z_{\mathcal{I}}(C)$ on the unit sphere of $V_{\mathcal{I}}$, the inclusion

$$\Big\{\frac{m}{|m|} \,:\, m \in M_{\epsilon,\mathcal{I}}\Big\} \subset U$$

holds for any positive ϵ small enough. Then, we can write $p = \cos\theta z + \sin\theta q$ for some q in $V_{\mathcal{I}}$ of norm 1 and orthogonal to z, and some θ close to 0 mod π — the coefficients $\cos\theta$ and $\sin\theta$ are imposed by $|p| = 1$; speaking geometrically, we parameterize the sphere in normal coordinates. Since $\mathrm{Proj}_{V_{\mathcal{I}}}(C + C^{\mathrm{T}})z = 0$ thanks to Lemma 8.2.23, equation (8.2.21) forces

$$\sin^2\theta\,\langle Cq, q\rangle = -\epsilon + \frac{1}{\lambda}\,. \tag{8.2.22}$$

If $\theta = 0 \bmod \pi$, this forces $\lambda = 1/\epsilon$ and $p = z$. Thus $m = z/\sqrt{\epsilon}$. Assume that $\theta \neq 0 \bmod \pi$. Since $S_{\mathcal{I}}(C)$ is invariant under $\{ -\mathrm{Id}, \mathrm{Id} \}$, equality (8.2.22) imposes

$$q = \frac{1}{\sin\theta}\sqrt{\epsilon - \frac{1}{\lambda}}\, v \qquad \text{for some } v \in S_{\mathcal{I}}(C)\,.$$

We can then write

$$m = \sqrt{\lambda} p = \sqrt{\lambda}\cos\theta\, z + \sqrt{\epsilon\lambda - 1}\, v\,.$$

But equation (8.2.20) yields

$$1 = -\epsilon\lambda + 1 + \epsilon(\lambda\cos^2\theta + \epsilon\lambda - 1)\,,$$

from which we can obtain $\cos^2\theta$. Thus,

$$m = \eta\sqrt{1 + \lambda(1-\epsilon)}\, z + \sqrt{\epsilon\lambda - 1}\, v \qquad \text{for } \eta \in \{ -1, 1 \}\,.$$

Set $s = \epsilon\lambda - 1$ to obtain the parameterization given in the lemma. Notice that for $s = -1$, we obtain $m = \eta\sqrt{\lambda} z$. ∎

We can now obtain some good bounds on the components of points m belonging to $M_{\epsilon,\mathcal{I}}$. We write $m_i = \langle m, e_i \rangle$ the i-th component of m, as we did in the proof of Theorem 8.2.10. The following statement is the analogue of Lemma 8.2.16.

8.2.25. LEMMA. *There exists a positive M such that for any positive ϵ small enough, any point $m = \frac{\eta}{\sqrt{\epsilon}}\sqrt{1 + s(1-\epsilon)}\, z + \sqrt{s}\, v$ in $M_{\epsilon,\mathcal{I}}$ satisfies for all i in $\mathcal{I}$ the inequality*

$$\frac{1}{M} \leqslant \sqrt{\frac{\epsilon}{1+s}}|m_i| \leqslant M\,.$$

Proof. For ϵ small enough and any positive s, we have $s/(2\epsilon) \leqslant s(1-\epsilon)/\epsilon \leqslant s/\epsilon$. If $\eta\langle z, e_i \rangle$ is positive, this implies

$$\sqrt{\frac{\epsilon}{1+s}} m_i \leqslant \eta\langle z, e_i \rangle + \sqrt{\frac{s\epsilon}{1+s}}\langle v, e_i \rangle$$

and

$$\sqrt{\frac{\epsilon}{1+s}} m_i \geqslant \sqrt{\frac{1 + (s/2)}{1+s}}\eta\langle z, e_i \rangle + \sqrt{\frac{s\epsilon}{1+s}}\langle v, e_i \rangle\,.$$

Since $S_{\mathcal{I}}(C)$ is compact and the functions $s \mapsto s/(1+s)$ and $s \mapsto (2+s)/(1+s)$ are bounded on $[0,\infty)$, the result follows.

If $\eta\langle z, e_i\rangle$ is negative, we proceed similarly. ■

We can conclude the proof of Theorem 8.2.21 with the next result.

8.2.26. LEMMA. *For any $\mathcal{I}$ in $J_0(C)$,*

$$\lim_{\epsilon\to 0}\int_{M_{\epsilon,\mathcal{I}}} \frac{\det \tilde{G}(m)}{\prod_{i\in\mathcal{I}} |m_i|^\alpha} \mathrm{d}\mathcal{M}_{M_{\epsilon,\mathcal{I}}}(m) = 0\,.$$

Proof. Arguing as in Lemma 8.2.17, we obtain

$$\det \tilde{G}(m) \leqslant M/|m|^{N_0-1}$$

for any m in $M_{\epsilon,\mathcal{I}}$ and some fixed positive M, not depending on ϵ or m. Using Lemma 8.2.25, we have

$$\begin{aligned} &\int_{M_{\epsilon,\mathcal{I}}} \frac{\det \tilde{G}(m)}{\prod_{i\in\mathcal{I}} |m_i|^\alpha} \mathrm{d}\mathcal{M}_{M_{\epsilon,\mathcal{I}}}(m) \\ &\qquad \leqslant \int_{M_{\epsilon,\mathcal{I}}} \Big(\frac{\epsilon}{1+s}\Big)^{\frac{N_0-1+N_0\alpha}{2}} \mathrm{d}\mathcal{M}_{M_{\epsilon,\mathcal{I}}}(m) O(1)\,, \end{aligned} \tag{8.2.22}$$

as ϵ tends to 0. To express the Riemannian measure on $M_{\epsilon,\mathcal{I}}$, notice that $S_{\mathcal{I}}(C)$ is a manifold of dimension N_0-2. Write $v(x_1,\ldots,x_{N_0-2})$ for a local parameterization of $S_{\mathcal{I}}(C)$. It induces a parameterization

$$m(s,x_1,\ldots,x_{N_0-2}) = \frac{\eta}{\sqrt{\epsilon}}\sqrt{1+s(1-\epsilon)}\,z + \sqrt{s}\,v(x_1,\ldots,x_{N_0-2})$$

of $M_{\epsilon,\mathcal{I}}$. In this local chart

$$\begin{aligned} \mathrm{d}\mathcal{M}_{M_{\epsilon,\mathcal{I}}}(m) = \Big| &\frac{\sqrt{1-\epsilon}\, s^{(N_0-2)/2}}{\sqrt{\epsilon}\sqrt{1+s(1-\epsilon)}} \det\Big(z, \frac{\partial v}{\partial x_1},\ldots,\frac{\partial v}{\partial x_{N_0-2}}\Big) \\ &+ \frac{s^{(N_0-2)/2}}{2\sqrt{s}} \det\Big(v, \frac{\partial v}{\partial x_1},\ldots,\frac{\partial v}{\partial x_{N_0-2}}\Big)\Big| \,\mathrm{d}s\wedge \mathrm{d}x_1 \wedge\ldots\wedge \mathrm{d}x_{N_0-2}\,. \end{aligned}$$

Since z is orthogonal to $S_{\mathcal{I}}(C)$ and of unit norm, we have

$$\det\Big(z, \frac{\partial v}{\partial x_1},\ldots,\frac{\partial v}{\partial x_{N_0-2}}\Big)\mathrm{d}x_1\wedge\ldots\wedge \mathrm{d}x_{N_0-2} = \mathrm{d}\mathcal{M}_{S_{\mathcal{I}}(C)}\,.$$

Therefore,

$$
\begin{aligned}
\mathrm{d}\mathcal{M}_{M_{\epsilon,\mathcal{I}}}(m) \leqslant & \sqrt{\frac{1-\epsilon}{1+s(1-\epsilon)}}\frac{1}{\sqrt{\epsilon}}\, s^{(N_0-2)/2}\,\mathrm{d}s\,\mathrm{d}\mathcal{M}_{S_{\mathcal{I}}(C)}(v) \\
&\qquad + \frac{s^{(N_0-3)/2}}{2}|\mathrm{Proj}_{(T_v S_{\mathcal{I}}(C))^\perp} v|\,\mathrm{d}s\,\mathrm{d}\mathcal{M}_{S_{\mathcal{I}}(C)}(v)\,.
\end{aligned}
$$

Consequently, (8.2.21) is less than

$$
\begin{aligned}
\int_{s>0}\int_{S_{\mathcal{I}}(C)} & \frac{\epsilon^{(N_0-2+N_0\alpha)/2}}{(1+s)^{(N_0-1+N_0\alpha)/2}}\frac{s^{(N_0-2)/2}}{\sqrt{1+s(1-\epsilon)}} \\
&\quad + \frac{\epsilon^{(N_0-1+N_0)/2}}{(1+s)^{(N_0-1+N_0\alpha)/2}}\frac{s^{(N_0-3)/2}}{2}|v|\,\mathrm{d}s\,\mathrm{d}\mathcal{M}_{S_{\mathcal{I}}(C)}(v)\,.
\end{aligned}
$$

We can bound the integral in v by the constant $\mathrm{diam} S_{\mathcal{I}}(C)\mathrm{Vol} S_{\mathcal{I}}(C)$, for instance.

Then, the integral in s is less than

$$
\begin{aligned}
2+2\int_1^\infty & s^{\frac{N_0-2}{2}-\frac{N_0-1+N_0\alpha}{2}-\frac{1}{2}} + s^{\frac{N_0-3}{2}-\frac{N_0-1+N_0\alpha}{2}}\,\mathrm{d}s \\
&= 2+4\int_1^\infty s^{-1-\frac{N_0\alpha}{2}}\mathrm{d}s < \infty\,.
\end{aligned}
$$

Therefore, the right hand side of (8.2.21) is less than a constant times

$$
\epsilon^{\frac{N_0(1+\alpha)}{2}-1} + \epsilon^{\frac{N_0-1+N_0\alpha}{2}}\,.
$$

This tends to 0 as ϵ tends to 0, provided α is strictly larger than $(2/N_0)-1$. This concludes the proof of Lemma 8.2.26 as well as that of Theorem 8.2.21. ∎

8.3. Heavy tail and degeneracy.

The results of section 8.2 are incomplete. The tail behavior of $\langle CX, X\rangle$ is not described when the largest diagonal coefficient of C vanishes, or when (8.2.8) does not hold. The proof of Theorem 8.2.10 breaks down in these cases.

In example 4 of chapter 1, we dealt with the tail of the product of two independent Cauchy random variables. In the language of the current chapter, we considered the 2×2-matrix

$$
C = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,.
$$

We saw that the tail of $\langle CX,X\rangle$ is in $t^{-1}\log t$. For this specific matrix $N(C)=2$. Comparing with the result obtained in the previous section, we have an extra logarithmic factor in t. Going back to chapter 1, the factor t^{-1} is explained by the fact that for most points (x,y) on the boundary of

$$A_t=\{\,(x,y)\,:\,xy\geqslant t\,\}=\sqrt{t}A_1\,,$$

the minimum of x and y tends to infinity, sharing some similarity with the case where $N(C)=2$ and (8.2.8) holds. The factor $\log t$ is explained by the closeness of the boundary ∂A_t to the level sets of the Cauchy density, along a sizeable part of the hyperbola $xy=t$.

Quite amazingly, Theorem 5.1 can still be used in this situation, but the analysis is a bit more involved than needed to prove Theorems 8.2.1 or 8.2.10. Our goal in this section is rather modest. We do not intend to obtain the tail behavior of the quadratic form whenever (8.2.8) does not hold. We will concentrate only on matrices C with vanishing largest diagonal coefficient. The reason is twofold. First, this simple degeneracy is sufficient to understand what we should do when (8.2.8) does not hold. Second, we will make use of this specific case in chapter 11, when studying a statistical application.

8.3.1. THEOREM. *Let C be a $d\times d$-real matrix, with vanishing largest diagonal coefficient. Let X be a random vector in $\mathbb{R}^d$, with independent components, having a Student-like distribution S_α. The following asymptotic equivalence holds as t tends to infinity,*

$$P\{\,\langle CX,X\rangle\geqslant t\,\}\sim\frac{\log t}{t^\alpha}K_{s,\alpha}^2\alpha^\alpha\sum_{i:C_{i,i}=0}\,\sum_{1\leqslant j\leqslant d}|C_{i,j}+C_{j,i}|^\alpha\,.$$

Proof. Since $\langle CX,X\rangle=\langle (C+C^{\mathrm{T}})X,X\rangle/2$, we will assume throughout the proof that C is symmetric. The proof builds upon that of Theorem 8.2.10 in the case $N(C)=2$. However, Lemma 8.2.11 cannot be used anymore since (8.2.8) does not hold.

As before, we denote

$$A_t=\{\,p\in\mathbb{R}^d\,:\,\langle Cp,p\rangle\geqslant t\,\}=\sqrt{t}A_1$$

and

$$B_t=\Phi^{\leftarrow}\circ S_\alpha(A_t)=\{\,\Phi^{\leftarrow}\circ S_\alpha(p)\,:\,p\in A_t\,\}\,.$$

This guarantees

$$P\{\langle CX, X\rangle \geqslant t\} = \frac{1}{(2\pi)^{d/2}} \int_{B_t} e^{-|y|^2/2} \mathrm{d}y\,.$$

The main part of the proof is to precisely locate and describe the points in B_t with nearly minimal norm. This will be done through the next five lemmas.

Since all the diagonal coefficients of C are nonpositive, $N(C) \geqslant 2$. Our first lemma shows that $N(C) = 2$, an easy fact. It also introduces points which will be essential to determine a dominating manifold. Define

$$J_*(C) = \big\{ \{ i, j \} \,:\, 1 \leqslant i, j \leqslant d\,,\, C_{i,i}C_{j,j} = 0 \text{ and } C_{i,j} \neq 0 \big\}\,.$$

If $\{ i, j \} \in J_*(C)$ then either $C_{i,i}$ or $C_{j,j}$ is null.

8.3.2. LEMMA. *Let $\mathcal{I} = \{ i, j \}$ be in $J_*(C)$, and assume that $C_{i,i} = 0$. The solutions of the equation $\langle Cm, m\rangle = 1$ in $V_{\mathcal{I}}$ are*

$$\frac{1}{2C_{i,j}}\Big(\frac{1}{m_j} - m_j C_{j,j}\Big) e_i + m_j e_j\,, \qquad m_j \in \mathbb{R} \setminus \{ 0 \}\,.$$

Proof. Let $m = m_i e_i + m_j e_j$ be a point in $V_{\mathcal{I}}$. The equation $1 = \langle Cm, m\rangle$ can be rewritten as

$$1 = 2C_{i,j} m_i m_j + C_{j,j} m_j^2\,.$$

Since $(m_i, 0)$ is not solution, we obtain the result in expressing m_i as a function of m_j. ∎

It follows from Lemma 8.3.2 that $J(C)$ contains $J_*(C)$. As we did in the proof of Theorem 8.2.10, define

$$M_{\mathcal{I}} = \{ m \in V_{\mathcal{I}} \,:\, \langle Cm, m\rangle = 1 \}\,, \qquad \mathcal{I} \in J(C)\,,$$

and

$$\gamma_{\mathcal{I}} = \inf \Big\{ \prod_{i \in \mathcal{I}} |m_i| \,:\, m \in M_{\mathcal{I}} \Big\}\,.$$

It will be convenient later to extend slightly the notation introduced so far. For $\mathcal{I} = \{ i, j \}$ in $J(C)$, we will write $M_{i,j}$ or $M_{j,i}$ for $M_{\mathcal{I}}$.

If $\mathcal{I}$ belongs to $J(C)\setminus J_*(C)$, Lemma 8.2.11 shows that $\gamma_{\mathcal{I}}$ is positive. If $\mathcal{I}=\{\,i,j\,\}$ is in $J_*(C)$, Lemma 8.3.2 gives

$$\gamma_{\mathcal{I}}=\inf\Bigl\{\,\frac{|1-m_j^2C_{j,j}|}{2|C_{i,j}|}\;:\;m_j\in\mathbb{R}\setminus\{\,0\,\}\,\Bigr\}=\frac{1}{2|C_{i,j}|}\,,$$

because $C_{j,j}$ is nonpositive. Consequently,

$$\gamma=\min\{\,\gamma_{\mathcal{I}}\;:\;\mathcal{I}\in J(C)\,\}$$

is positive. As in Theorem 8.2.10, we hope

$$I(y)=\frac{|y|^2}{2}-\log(2\pi)^{d/2}$$

will be minimum on B_t at points lying in $\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}\bigcup_{\mathcal{I}\in J(C)}M_{\mathcal{I}})$. Let

$$\begin{aligned}R(t)=2\alpha\log\sqrt{t}-\log\log\sqrt{t}-2\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})&\\+\log(2\pi)^{d/2}+\alpha\log\gamma\,.&\end{aligned}$$

If $\mathcal{I}$ and m are fixed, respectively in $J(C)$ and $M_{\mathcal{I}}$, the proof of Theorem 8.2.10 — see the calculation of $I\bigl(q_{\mathcal{I},t}(m,v)\bigr)$ before Lemma 8.2.13 — shows that

$$I\bigl(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}m)\bigr)=R(t)-\alpha\log\gamma+\alpha\log\Bigl(\prod_{i\in\mathcal{I}}|m_i|\Bigr)+o(1)$$

as t tends to infinity. This suggests that $I(B_t)=R(t)+o(1)$ as t tends to infinity. To locate the minima of I on B_t, we try to imitate the second assertion of Lemma 8.2.13. The coming lemma gives a coarser estimate. After its proof, we will be able to explain more precisely the difficulty created by the nonemptyness of $J_*(C)$.

8.3.3. LEMMA. *Let M be a number strictly larger than 2, and ϵ be positive. The set of all points p in ∂A_1 such that*

$$I\bigl(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\bigr)\leqslant R(t)+M\log\log t$$

is included in an $\epsilon t^{-1/6}(\log t)^{M/\alpha}$-neighborhood of $\bigcup_{\mathcal{I}\in J(C)}M_{\mathcal{I}}$.

Proof. For any point p, we use Lemma A.1.5 to obtain

$$\begin{aligned}
&I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\big)\\
&\quad\geqslant \frac{1}{2}\sum_{1\leqslant i\leqslant d}\Phi^{\leftarrow}\circ S_\alpha\big(\epsilon t^{\frac12-\frac16}(\log t)^{M/\alpha}\big)\mathrm{I}_{[\epsilon t^{-1/6}(\log t)^{M/\alpha},\infty)}(|p_i|)\\
&\quad=\frac12\Big(2\alpha\log\big(t^{1/3}(\log t)^{\frac{M}{\alpha}}\big)^2-\log\log\big(t^{1/3}(\log t)^{M/\alpha}\big)\big(1+o(1)\big)\Big)\times\\
&\qquad\qquad\sharp\big\{\,1\leqslant i\leqslant d\,:\,|p_i|\geqslant\epsilon t^{-1/6}(\log t)^{M/\alpha}\,\big\}\,.
\end{aligned}$$

Consequently, for all the points under consideration in the lemma, and for t large enough,

$$\begin{aligned}
\sharp\big\{\,1\leqslant i\leqslant d\,:\,|p_i|\geqslant\epsilon t^{-1/6}(\log t)^{M/\alpha}\,\big\}&\\
&\leqslant\frac{\alpha\log t+(M-1)\log\log t+O(1)}{\frac{\alpha}{3}\log t+\big(M-\frac12\big)\log\log t\big(1+o(1)\big)}\\
&\leqslant 3-\frac{3}{\alpha}(2M-4)\frac{\log\log t}{\log t}\big(1+o(1)\big)\\
&<3\,.
\end{aligned}$$

Thus, at most two of the p_i's have their absolute value larger than $\epsilon t^{-1/6}(\log t)^{M\alpha}$. Since $N(C)=2$, this concludes the proof. ∎

Now, let us see precisely why the proof of Theorem 8.2.10 breaks down. Lemma 8.2.11 still holds for $\mathcal{I}$ in $J(C)\setminus J_*(C)$. But it fails if $\mathcal{I}=\{\,i,j\,\}$ is in $J_*(C)$, since Lemma 8.3.2 shows that we can have $|m_i|$ as small as we like; and whenever $C_{j,j}$ vanishes, $|m_j|$ can also be as close to 0 as desired. Lemma 8.2.11 was used in deriving the expression of $q_{\mathcal{I},t}(m,v)$ in Lemma 8.2.12. Then all the proof was more or less based on this approximation of $q_{\mathcal{I},t}(m,v)$. Notice that we can still use this expression whenever we can make the expansion which was used in its proof. Thus, for $\mathcal{I}=\{\,i,j\,\}$ in $J(C)$, we still parameterize ∂A_1 near a point m of $M_{\mathcal{I}}$, as

$$p_{\mathcal{I}}(m,v)=m\big(1+h(v)\big)+v\,,\qquad v\in T_m\partial A_1\ominus T_mM_{\mathcal{I}}\,.$$

We need to be able to prove that only those v's such that $v\ll m$ componentwise are of interest for us. Thanks to Lemma 8.3.3, this can be done right away on the range where

$$|m_i|\wedge|m_j|\geqslant t^{-1/6}(\log t)^{M/\alpha}\tag{8.3.1}$$

say. But it can be seen easily that we need a larger range when $\mathcal{I}$ belongs to $J_*(C)$.

Another difference with the situation of Theorem 8.2.10 is that the sets $M_{\mathcal{I}}$ are no longer far apart. It is true that the distance between $\bigcup_{\mathcal{I}\in J_*(C)} M_{\mathcal{I}}$ and $\bigcup_{\mathcal{I}\in J(C)\setminus J_*(C)} M_{\mathcal{I}}$ is strictly positive. Also the distance between $M_{\mathcal{I}}$ and $M_{\mathcal{J}}$ for $\mathcal{I}$ and $\mathcal{J}$ distinct in $J(C)\setminus J_*(C)$ is positive. This follows from Lemma 8.2.11. However, if $\{\, i,j\,\}$ and $\{\, i,k\,\}$ are in $J_*(C)$ with $j\neq k$ then $M_{i,j}$ and $M_{i,k}$ are at zero distance. They connect at infinity on the axis $e_i\mathbb{R}$. This can be seen as follows. For $m_k = m_jC_{i,j}/C_{i,k}$, the corresponding point on $M_{i,k}$ given by Lemma 8.3.2 is

$$\begin{aligned}\frac{1}{2C_{i,k}}\Big(\frac{1}{m_k} - m_kC_{k,k}\Big)e_i + m_ke_k \\ = \Big(\frac{1}{2C_{i,j}m_j} - \frac{C_{i,j}C_{k,k}}{2C_{i,k}^2}m_j\Big)e_i + \frac{m_jC_{i,j}}{C_{i,k}}e_k\,.\end{aligned}$$

Its distance to

$$\frac{1}{2C_{i,j}}\Big(\frac{1}{m_j} - m_jC_{j,j}\Big)e_i + m_je_j \in M_{i,j}$$

is

$$\bigg(\Big(\frac{C_{i,j}C_{k,k}}{2C_{i,k}^2}\Big)^2 + \Big(\frac{C_{i,j}}{C_{i,k}} - 1\Big)^2\bigg)^{1/2}|m_j|\,.$$

It tends to 0 as m_j tends to 0. This has a dramatic consequence. For m_j small, we can have $p_{\{i,j\}}(m,v) = p_{\{i,k\}}(m',v')$ for some m in $M_{i,j}$, some m' in $M_{i,k}$ and v, v' satisfying the a priori estimate given in Lemma 8.3.2. In other words some components of v, even small, may cancel with the corresponding components of m. Thus, the naive approach used in Lemma 8.2.12 cannot succeed.

It is essential to understand that what goes wrong here is the parameterization of the set ∂A_1. The naive parameterization with (m,v) is onto, but is not one-to-one anymore, at least in the interesting range. The trick is then to introduce a new parameterization, well tailored to handle small components of the points m in $\bigcup_{\mathcal{I}\in J_*(C)} M_{\mathcal{I}}$. Lemma 8.3.2 shows that when (8.3.1) fails, then one component of m has to be of order larger than $t^{1/6}(\log t)^{M/\alpha}$. The corresponding component of v in $p_{\mathcal{I}}(m,v)$ can be neglected thanks to Lemma 8.3.3. Writing p for the corresponding point in ∂A_1, we have $p_i\sim m_i$. Since p_i is very large and Lemma 8.3.3 tells us to look in a neighborhood of m, all the

other components of p must be small, going to 0 as t tends to infinity. We can then single out another component of p, say p_k, such that we can locate p near $M_{i,k}$. To do so, write

$$p = p_i e_i + p_k e_k + w$$

where w is orthogonal to $V_{\{i,k\}}$. Since p is in ∂A_1 and $C_{i,i}$ vanishes,

$$1 = 2C_{i,k}p_ip_k + p_k^2 C_{k,k} + 2p_i\langle Ce_i, w\rangle + 2p_k\langle Ce_j, w\rangle + \langle Cw, w\rangle\,.$$

Using Lemma 8.3.2, we focus on a zone where p_k and w are $o(1)$ as t tends to infinity, and obtain

$$1 = 2C_{i,k}p_ip_k + 2p_i\langle Ce_i, w\rangle + o(1)\,.$$

Notice that

$$Ce_i \in \operatorname{span}\bigl\{\, e_l \,:\, \{\, i,l\,\} \in J_*(C)\,\bigr\}\,.$$

Thus, whenever $\{\, i,l\,\}$ is in $J_*(C)$, if we can guarantee $|w_l| \leqslant |p_k|$, we obtain

$$1 \leqslant (2|C_{i,k}| + \|C\|\sqrt{d})|p_ip_k| + o(1)\,.$$

This gives us a lower bound on $|p_k|$. It will be good enough to bound $\sqrt{t}p_k$ away from 0, and to do an asymptotic expansion of $\Phi^{\leftarrow}\circ S_\alpha\bigl(\sqrt{t}p_k)\bigr)^2$ in the calculation of $I\bigl(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\bigr)$. We will then be able to improve the a priori bound of Lemma 8.3.3. Only then we will use the parameterization $p_{\mathcal{I}}(m,v)$.

We have not said how to guarantee that $|w_l| \leqslant |p_k|$. This is easy. Take p_k to be the largest of the $|p_j|$'s for $\{\, i,j\,\}$ in $J_*(C)$. This index k depends on $p\in\partial A_1$, and this parameterization is one-to-one up to a set of Lebesgue measure zero.

To proceed along these lines, define the function

$$Q(x) = \sqrt{2\alpha\log\sqrt{x}} - \frac{\log\log\sqrt{x}}{2\sqrt{2\alpha\log\sqrt{x}}} - \frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{\sqrt{2\alpha\log\sqrt{x}}}\,, \qquad x>0\,.$$

8.3.4. LEMMA. *Let $\mathcal{I} = \{\, i,j\,\}$ be in $J(C)$. Uniformly in the range m in $M_{\mathcal{I}}$ and v in $T_m\partial A_1 \setminus T_mM_{\mathcal{I}}$ such that $\sqrt{t}(|m_i|\wedge|m_j|)$ tends to infinity, $v_i = o(m_i)$ and $v_j = o(m_j)$,*

$$\Phi^{\leftarrow}\circ S_\alpha\bigl(\sqrt{t}p_{\mathcal{I}}(m,v)\bigr) = \sum_{k\in\mathcal{I}} \operatorname{sign}(m_k)\Bigl(Q(tm_k^2) + o\bigl(\log(tm_k^2)\bigr)^{-1/2}\Bigr)e_k + \Phi^{\leftarrow}\circ S_\alpha(\operatorname{Proj}_{V_{\mathcal{I}}^{\perp}}v)\,.$$

Proof. In the given range, for $k = i, j$,

$$\langle \sqrt{t} p_{\mathcal{I}}(m,v), e_k \rangle = \sqrt{t} m_k \bigl(1 + o(1)\bigr)$$

tends to infinity. We can apply Lemma A.1.5 to obtain an asymptotic expansion of $\Phi^{\leftarrow} \circ S_\alpha\bigl(\langle \sqrt{t} p_{\mathcal{I}}(m,v), e_k\rangle\bigr)$. This gives the result. ∎

In particular, for the points considered in Lemma 8.3.4, as t tends to infinity,

$$\begin{aligned} I\Bigl(\Phi^{\leftarrow} \circ S_\alpha\bigl(\sqrt{t} p_{\mathcal{I}}(m,v)\bigr)\Bigr) & \\ = \sum_{k \in \mathcal{I}} & \Bigl(\alpha \log\bigl(\sqrt{t}|m_k|\bigr) - \frac{1}{2} \log\log\bigl(\sqrt{t}|m_k|\bigr)\Bigr) \\ & - 2\log(K_{s,\alpha} \alpha^{\alpha/2} 2\sqrt{\pi}) + \log(2\pi)^{d/2} \\ & + \frac{1}{2} |\Phi^{\leftarrow} \circ S_\alpha(\mathrm{Proj}_{V_{\mathcal{I}}^{\perp}} v)|^2 + o(1)\,. \end{aligned} \tag{8.3.2}$$

As we explained above, Lemmas 8.3.3 and 8.3.4 give us a suitable approximation of ∂B_t near points parametrized by $m = m_i e_i + m_j e_j$ in $M_{i,j}$ with $|m_i| \wedge |m_j| \geqslant t^{-1/6}(\log t)^{M/\alpha}$. Those points have two components greater than $t^{-1/6}(\log t)^{M/\alpha}\bigl(1+o(1)\bigr)$. If now a point p in ∂A_1 has only one component larger than $t^{-1/6}(\log t)^{M/\alpha}$ and is in a $o(1)$-neighborhood of $\bigcup_{\mathcal{I} \in J(C)} M_{\mathcal{I}}$, then one component, say p_i, has to blow up with t. Then, p has to be in a $o(1)$-neighborhood of $\bigcup_{k : C_{i,k} \neq 0} M_{i,k}$. The following lemma gives some control on the largest component of p_k for $C_{i,k}$ nonzero.

8.3.5. LEMMA. *Let ϵ be a positive number. Let i be such that $C_{i,i}$ is null. Furthermore, let p be a point in ∂A_1 and in a $o(1)$-neighborhood of $\bigcup_{k : C_{i,k} \neq 0} M_{i,k}$, as t tends to infinity, with $|p_i| \geqslant \epsilon$. Choose j such that*

$$|p_j| = \max\{\, |p_k| \,:\, C_{i,k} \neq 0 \,\}\,.$$

Then, for t large enough,

$$|p_i p_j| \geqslant \frac{1}{8(d\|C\| + \epsilon^{-1})}\,.$$

In particular, for t large enough, the inequality $|p_i| \leqslant r$ forces

$$|p_j| \geqslant \frac{1}{8r(d\|C\| + \epsilon^{-1})}\,.$$

Proof. Since $\bigcup_{k:C_{i,k}\neq 0} M_{i,k}$ is included in the union of planes generated by (e_i, e_k) for $C_{i,k}$ nonzero, at most two components of p are not $o(1)$. Moreover, one component is p_i and the other one is a p_k for $C_{i,k}$ nonzero, that is p_j. Write $p = p_i e_i + p_j e_j + w$ with w orthogonal to both e_i and e_j. By definition of p_j, each component of w is smaller than $|p_j|$. From what precedes, the norm of w has to be less than $\sqrt{d}\,|p_j|$, as well as $o(1)$. Since p belongs to ∂A_1 and Ce_i is in $\mathrm{span}\{\, e_k \,:\, C_{i,k}\neq 0\,\}$,

$$\begin{aligned} 1 = \langle Cp, p\rangle &= 2p_ip_jC_{i,j} + 2p_i\langle Ce_i, w\rangle + 2p_j\langle Ce_j, w\rangle + \langle Cw, w\rangle \\ &\leqslant 2|p_ip_j|\|C\| + 2\sqrt{d}\,|p_ip_j|\|C\| + |p_j|o(1) + o(1) \end{aligned}$$

as t tends to infinity. Since $|p_i|\geqslant \epsilon$, we obtain

$$1 \leqslant |p_ip_j|\big(2(1+\sqrt{d})\|C\| + 4\epsilon^{-1}\big) + o(1)\,.$$

This implies the first statement of the lemma. The second follows trivially. ∎

The lower bounds in Lemma 8.3.5 are useful only if p_i and p_j are not too large, so that we can have $\sqrt{t}(|p_i|\wedge|p_j|)$ going to infinity. Thus we need to shrink the domain on which we need to perform the integration. This relies on the simple observation that for Student-like distributions, a Bonferoni type inequality gives

$$P\Big\{ \max_{1\leqslant i\leqslant d} |X_i| \geqslant \frac{tM}{(\log t)^{1/\alpha}} \Big\} \leqslant 2d\frac{\log t}{t^\alpha M^\alpha} K_{s,\alpha}\alpha^{(\alpha-1)/2} \qquad (8.3.3)$$

for t large enough. Thus, we define the subset

$$A'_t = \sqrt{t}A'_1 = \sqrt{t}\,\Big\{\, p\in\mathbb{R}^d \,:\, \langle Cp, p\rangle \geqslant 1\,,\ \max_{1\leqslant i\leqslant d}|p_i| \leqslant \frac{\sqrt{t}\log\log t}{(\log t)^{1/\alpha}} \,\Big\}\,.$$

We denote $B'_t = \Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}A'_t)$.

We can now improve dramatically upon Lemma 8.3.3.

8.3.6. LEMMA. *The set of all p's in $\partial A'_1$ such that*

$$I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\big) \leqslant R(t) + M\log\log t$$

can be parameterized as $p(m,v)$ with m in some $M_{\mathcal{I}}$, some $\mathcal{I}$ in $J(C)$ and

$$\mathrm{Proj}_{V_{\mathcal{I}}^\perp} v = O\big(t^{-1/2}(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}\big)\,.$$

Moreover, whenever $|m_i| \wedge |m_j| \gg t^{-1/2}(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}$, *then* $v_i = o(m_i)$ *and* $v_j = o(m_j)$. *In addition,*

$$I\big(\Phi^{\leftarrow} \circ S_\alpha(\sqrt{t}A'_1)\big) \geqslant R(t) - \alpha\log\gamma - \alpha\log(16\,d\,\|C\|) + o(1)$$

as t *tends to infinity.*

Proof. Let p be a point in $\partial A'_1$ as in the statement of the lemma. Lemma 8.3.3 guarantees that it belongs to an $o(t^{-1/6}(\log)^{M/\alpha})$-neighborhood of $\bigcup_{\mathcal{I}\in J(C)} M_{\mathcal{I}}$. If it is in an $o(t^{-1/6}(\log t)^{M/\alpha})$-neighborhood of $\bigcup_{\mathcal{I}\in J(C)\setminus J_*(C)} M_{\mathcal{I}}$, then we are done thanks to Lemmas 8.3.4 and 8.2.11. Thus, assume that it is not. Let i be such that

$$|p_i| = \max\{\,|p_k| \,:\, 1 \leqslant k \leqslant d\,\}\,.$$

Since $1 = \langle Cp, p\rangle$ is less than $\|C\||p|^2$, we have

$$|p_i| \geqslant 1/d\|C\| \gg t^{-1/6}(\log t)^{M/\alpha}\,.$$

Therefore, p is in a $o\big(t^{-1/6}(\log t)^{M/\alpha}\big)$-neighborhood of $\bigcup_{k:C_{i,k}\neq 0} M_{i,k}$. Since p belongs to A'_1,

$$|p_i| \leqslant \frac{\sqrt{t}\log\log t}{(\log t)^{1/\alpha}}\,.$$

Let p_j be as in Lemma 8.3.5. The second inequality in Lemma 8.3.5 gives

$$|p_j| \geqslant \frac{(\log t)^{1/\alpha}}{\sqrt{t}\log\log t}\,\frac{1}{16\,d\,\|C\|}\,.$$

In particular, both $\sqrt{t}|p_i|$ and $\sqrt{t}|p_j|$ tend to infinity. Write $p = p_ie_i + p_je_j + w$ with w orthogonal to both e_i and e_j. We calculate

$$\begin{aligned}
I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\big) &= \frac{1}{2}|\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p_i)|^2 + \frac{1}{2}|\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p_j)|^2\\
&\qquad + \frac{1}{2}|\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}w)|^2 + \log(2\pi)^{d/2}\\
&= \alpha\log(\sqrt{t}|p_i|) - \frac{1}{2}\log\log(\sqrt{t}|p_i|) + \alpha\log(\sqrt{t}|p_j|)\\
&\qquad - \frac{1}{2}\log\log(\sqrt{t}|p_j|) - 2\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})\\
&\qquad + \frac{1}{2}|\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}w)|^2 + \log(2\pi)^{d/2} + o(1)
\end{aligned}$$

as t tends to infinity. From the first estimate in Lemma 8.3.5, we deduce

$$\alpha\log(\sqrt{t}|p_i|)+\alpha\log(\sqrt{t}|p_j|)\geqslant 2\alpha\log\sqrt{t}-\alpha\log(16\,d\,\|C\|)\,.$$

Moreover, since $\log(1+h)\leqslant h$ for any $h>-1$, we also have

$$\log\log(\sqrt{t}|p_i|)=\log(\log\sqrt{t}+\log|p_i|)\leqslant\log\log\sqrt{t}+\frac{\log|p_i|}{\log\sqrt{t}}\,.$$

Thus,

$$\begin{aligned}I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\big)\geqslant{}&2\alpha\log\sqrt{t}-\log\log\sqrt{t}-2\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})\\&+\log(2\pi)^{d/2}-\alpha\log(16d\|C\|)+o(1)\\&+\frac{1}{2}|\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}w)|^2\,,\end{aligned}$$

where the $o(1)$-term does not depend on p in A_1'. This gives the lower bound on $I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}A_1')\big)$ stated in the Lemma. Moreover, the inequality

$$I\big(\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}p)\big)\leqslant R(t)+M\log\log t$$

implies

$$|\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}w)|^2\leqslant M\log\log t+c$$

for some constant c. This forces, for any k not in $\{\,i,j\,\}$,

$$\sqrt{t}|w_k|\leqslant S_\alpha^{\leftarrow}\circ\phi(\sqrt{M\log\log t+c})=(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}O(1)$$

as t tends to infinity, thanks to Lemma A.1.6.

Set $\mathcal{I}=\{\,i,j\,\}$. We just showed that the projection of p on $V_{\mathcal{I}}^{\perp}$ is at most of order $t^{-1/2}(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}$ as t tends to infinity. From Lemma 8.3.3, we then know that $p=p(m,v)$ for some m in $M_{\mathcal{I}}$ and v in $T_m\partial A_1\setminus T_mM_{\mathcal{I}}$ with $|\mathrm{Proj}_{V_{\mathcal{I}}}v|=o\big(t^{-1/6}(\log t)^{M/\alpha}\big)$. Therefore, if $|m_i|$ and $|m_j|$ are at least $t^{-1/6}(\log t)^{M/\alpha}$, the lemma is proved. Our choice of i implies that the only case left to investigate is $|m_j|\leqslant t^{-1/6}(\log t)^{M/\alpha}$. In this case, Lemma 8.3.2 shows that $C_{i,i}$ must vanish — because m_i is lower bounded if $C_{j,j}$ is not null in Lemma 8.3.2; this is what we are claiming up to a permutation of i and j. Then, Lemma 8.3.2 shows that $m_i\sim 1/(2C_{i,j}m_j)$ as t tends to infinity. Since v is in $T_m\partial A_1\ominus T_mM_{\mathcal{I}}$, it is orthogonal to the vector spanning $T_mM_{\mathcal{I}}$, that is, to

$$\frac{1}{2C_{i,j}}\Big(-\frac{1}{m_j^2}-C_{j,j}\Big)e_i+e_j\,.$$

Thus, there exists a real r such that

$$\mathrm{Proj}_{V_{\mathcal{I}}} v = re_i + \frac{r}{2C_{i,j}}\Big(\frac{1}{m_j^2} + C_{j,j}\Big)e_j\,. \tag{8.3.4}$$

Since v belongs to $T_m\partial A_1$, we must have $\langle Cm, v\rangle = 0$, or, equivalently,

$$\langle \mathrm{Proj}_{V_{\mathcal{I}}} Cm, \mathrm{Proj}_{V_{\mathcal{I}}} v\rangle = \langle Cm, \mathrm{Proj}_{V_{\mathcal{I}}} v\rangle = -\langle Cm, \mathrm{Proj}_{V_{\mathcal{I}}^\perp} v\rangle\,. \tag{8.3.5}$$

Since $C_{i,i}$ is null,

$$\begin{aligned}\langle \mathrm{Proj}_{V_{\mathcal{I}}} Cm, \mathrm{Proj}_{V_{\mathcal{I}}} v\rangle &= \Big\langle \frac{1}{2}\Big(\frac{1}{m_j} - m_jC_{j,j}\Big)e_j + m_jC_{i,j}e_i + m_jC_{j,j}e_j\,, \\ &\qquad\qquad re_i + \frac{r}{2C_{i,j}}\Big(\frac{1}{m_j^2} + C_{j,j}\Big)e_j\Big\rangle \\ &= r_jC_{i,j} + \frac{r}{4C_{i,j}}\Big(\frac{1}{m_j} + m_jC_{j,j}\Big)\Big(\frac{1}{m_j^2} + C_{j,j}\Big)\,.\end{aligned}$$

Thus, if m_j tends to 0 as t tends to infinity,

$$\langle \mathrm{Proj}_{V_{\mathcal{I}}} Cm, \mathrm{Proj}_{V_{\mathcal{I}}} v\rangle \sim \frac{r}{4C_{i,j}m_j^3}\,.$$

On the other hand, using the bound on the projection of p on $V_{\mathcal{I}}^\perp$ that we obtained earlier in the proof of this lemma,

$$\begin{aligned}|\langle Cm, \mathrm{Proj}_{V_{\mathcal{I}}^\perp} v\rangle| &\leqslant \|C\|\,|m|\,|\mathrm{Proj}_{V_{\mathcal{I}}^\perp}(v)| \\ &= \frac{1}{|m_j|}O\big(t^{-1/2}(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}\big)\,.\end{aligned}$$

Therefore, (8.3.5) yields

$$r = m_j^2 O\big(t^{-1/2}(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}\big)\,.$$

as t tends to infinity. Therefore, going back to (8.3.4),

$$\mathrm{Proj}_{V_{\mathcal{I}}} v = o(m_i)e_i + \Big(O\big(t^{-1/2}(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}\big) + o(m_j)\Big)e_j.$$

This concludes the proof of Lemma 8.3.6. ∎

Though we do not need this right now, Lemmas 8.3.3 and 8.3.6 imply

$$I(B_t') \sim R(t) \sim \alpha\log t \qquad \text{as } t\to\infty\,.$$

More importantly, if i belongs to $\mathcal{I} = \{\, i, j \,\} \in J_*(C)$, and m is in $M_{\mathcal{I}}$, Lemma 8.3.2 shows that $m_j \sim 1/2C_{i,j}m_i$ as $|m_i|$ tends to infinity. Therefore, Lemma 8.3.6 allows us to use the expansion of Lemma 8.3.4 in the range $|m_i| \ll \sqrt{t}/(\log t)^{M/(2\alpha)}(\log\log t)^{1/(2\alpha)}$.

To conclude what we have done so far, we can use Lemma 8.3.4 over all points $q = \Phi^{\leftarrow} \circ S_\alpha(\sqrt{t}p)$ in B_t, with p belonging to A'_t, except for those in an $O\big(t^{-1/6}(\log t)^{M/\alpha}\big)$-neighborhood of a point $m = m_ie_i + m_je_j$ with $\{\, i, j \,\}$ in $J_*(C)$ and, say,

$$\sqrt{t}/(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha} \leqslant |m_i| \vee |m_j| \leqslant M\sqrt{t}/(\log t)^{1/\alpha}\,.$$

We now discard this missing range by an ad hoc argument. The key observation is that Lemma 8.3.5 guarantees that this set of p's is included in

$$\begin{aligned}\Omega_t = \bigcup_{\{i,j\}\in J_*(C)} \Big\{\, p \in \mathbb{R}^d \,:\, \frac{\sqrt{t}}{(\log t)^{M/\alpha}(\log\log t)^{1/(2\alpha)}} &\leqslant |p_i| \vee |p_j| \\ &\leqslant \frac{\sqrt{t}\log\log t}{(\log t)^{1/\alpha}}\,;\, |p_ip_j| \geqslant \frac{1}{16\,d\,\|C\|} \,\Big\}\,.\end{aligned}$$

We now show that ultimately we will be able to discard the set Ω_t in our computation.

8.3.7. LEMMA. *We have*

$$P\{\, X \in \sqrt{t}\Omega_t \,\} = o\Big(\frac{\log t}{t^\alpha}\Big) \qquad \textit{as } t \to \infty\,.$$

Proof. The upper and lower tail of the Student-like distributions are asymptotically equivalent. The result is then a consequence of the independence of the X_i's and the following calculation,

$$\begin{aligned} P\Big\{\, &\frac{t}{(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha}} \leqslant X_i \leqslant \frac{t\log\log t}{(\log t)^{1/\alpha}}\,;\, X_iX_j \geqslant \frac{t}{16d\|C\|} \,\Big\} \\ &= \int_{t(\log t)^{-M/(2\alpha)}(\log\log t)^{-1/\alpha}}^{t\log\log t(\log t)^{-1/\alpha}} 1 - S_\alpha(t/16d\|C\|x)\,\mathrm{d}S_\alpha(x) \\ &= O(1)\int_{t(\log t)^{-M/(2\alpha)}(\log\log t)^{-1/\alpha}}^{t\log\log t(\log t)^{-1/\alpha}} \frac{x^\alpha}{t^\alpha}\,\mathrm{d}S_\alpha(x)\,.\end{aligned}$$

An integration by parts shows that the last integral is

$$\frac{1}{t^\alpha}\big[x^\alpha\big(S_\alpha(x)-1\big)\big]_{t(\log t)^{-M/(2\alpha)}(\log\log t)^{-1/\alpha}}^{t\log\log t(\log t)^{-1/\alpha}}
+\frac{\alpha}{t^\alpha}\int_{t(\log t)^{-M/(2\alpha)}(\log\log t)^{-1/\alpha}}^{t\log\log t(\log t)^{-1/\alpha}} x^{\alpha-1}\big(1-S_\alpha(x)\big)\,\mathrm{d}x\,.$$

Since $x^\alpha\big(1-S_\alpha(x)\big)$ tends to a constant as x tends to infinity, the above sum is

$$\begin{aligned}o(t^{-\alpha})+O(1)t^{-\alpha}&\int_{t(\log t)^{-M/(2\alpha)}(\log\log t)^{-1/\alpha}}^{t\log\log t(\log t)^{-1/\alpha}} x^{-1}\mathrm{d}x\\ &=o(t^{-\alpha})+O(1)t^{-\alpha}\log\log t=o(t^{-\alpha}\log t)\end{aligned}$$

as t tends to infinity. ■

Define

$$B_t''=B_t'\setminus\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}\Omega_t)\,.$$

Combining (8.3.3) and Lemma 8.3.7, we obtain

$$P\{\,\langle CX,X\rangle\geqslant t\,\}=\int_{B_t''}e^{-I(y)}\mathrm{d}y+o\Big(\frac{\log t}{t^\alpha}\Big)\,.$$

From now on, we concentrate on the integral of e^{-I} on B_t''. Combining Lemmas 8.3.4, 8.3.5 and 8.3.6, the set $B_t''\bigcap\Gamma_{R(t)+M\log\log t}$ can be identified, for t large enough, with some points $\Phi^{\leftarrow}\circ S_\alpha\big(\sqrt{t}p_{\mathcal{I}}(m,v)\big)$, with

$$\begin{aligned}&\mathcal{I}\in J(C),\qquad m\in M_{\mathcal{I}},\qquad v\in T_m\partial A_1\setminus T_mM_{\mathcal{I}}\\ &|\mathrm{Proj}_{V_{\mathcal{I}}^\perp}|\leqslant t^{-1/2}(\log t)^{2M/\alpha}(\log\log t)^{1/(2\alpha)},\\ &\min\nolimits_{k\in\mathcal{I}}|m_k|\geqslant t^{-1/2}(\log\log t)^{1/\alpha}(\log t)^{M/(2\alpha)},\end{aligned}$$

and $\qquad\mathrm{Proj}_{V_{\mathcal{I}}}v=o(m)$ componentwise.

For those points, we can use Lemma 8.3.4 and (8.3.2). This allows us to obtain the value of $I(B_t'')$ up to an $o(1)$-term as t tends to infinity.

8.3.8. LEMMA. *The equality $I(B_t'')=R(t)+o(1)$ holds as t tends to infinity.*

Proof. Making use of formula (8.3.2) we first need to evaluate the minimum of the function

$$\sum_{k\in\mathcal{I}}\Big(\alpha\log(\sqrt{t}|m_k|)-\frac{1}{2}\log\log(\sqrt{t}|m_k|)\Big)$$

on $M_{\mathcal{I}}$. We also need an approximate location of the minimum, to check that $\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}\,\cdot)$ maps it to a point in B_t'' — and not in B_t. When $\mathcal{I}$ is in $J(C)\setminus J_*(C)$, we can argue as in (8.2.9). We concentrate on the case where $\mathcal{I}=\{\,i,j\,\}$ belongs to $J_*(C)$. Considering Lemma 8.3.2, this leads us to define the function

$$\begin{aligned}f_{t,C_{j,j}}(m_j)&=\alpha\log\Big(\frac{\sqrt{t}}{2|C_{i,j}|}\Big|\frac{1}{m_j}-m_jC_{j,j}\Big|\Big)+\alpha\log(\sqrt{t}|m_j|)\\&\quad-\frac{1}{2}\log\log\Big(\frac{\sqrt{t}}{2|C_{i,j}|}\Big|\frac{1}{m_j}-m_jC_{j,j}\Big|\Big)-\frac{1}{2}\log\log(\sqrt{t}|m_j|)\,.\end{aligned}$$

When $C_{j,j}$ vanishes, this function is

$$\begin{aligned}f_{t,0}(m_j)=\alpha\log t-\alpha\log(2|C_{i,j}|)-\frac{1}{2}\log\log\frac{\sqrt{t}}{|C_{i,j}m_j|}&\\-\frac{1}{2}\log\log(\sqrt{t}|m_j|)\,.&\end{aligned}$$

The function $f_{t,0}(\cdot)$ is minimum when $r=\log|m_j|$ maximizes

$$(\log\sqrt{t}-\log|C_{i,j}|-r)(\log\sqrt{t}+r)\,,$$

that is

$$r=-\frac{1}{2}\log|C_{i,j}|\,.$$

Moreover, as t tends to infinity,

$$f_{t,0}\Big(\frac{1}{\sqrt{|C_{i,j}|}}\Big)=\alpha\log t-\alpha\log(2|C_{i,j}|)-\log\log\sqrt{t}+o(1)\,.$$

When $m_j=1/\sqrt{|C_{i,j}|}$ and $C_{j,j}$ vanishes, Lemma 8.3.2 gives $m_i=\mathrm{sign}(C_{i,j})/(2\sqrt{|C_{i,j}|})$. Clearly $m_ie_i+m_je_j$ is mapped into B_t'' by $\Phi^{\leftarrow}\circ S_\alpha(\sqrt{t}\,\cdot)$.

If $C_{j,j}$ is nonzero and m_j is positive and fixed, then

$$\begin{aligned}f_{t,C_{j,j}}(m_j)=\alpha\log t-\alpha\log(2|C_{i,j}|)+\alpha\log(1-m_j^2C_{j,j})&\\-\log\log\sqrt{t}+o(1)&\qquad(8.3.6)\end{aligned}$$

— recall that $C_{j,j}$ is negative. Thus, for $f_{t,C_{j,j}}(m_j)$ to be minimum, we must have m_j^2 tends to 0 as t tends to infinity. But if $\sqrt{t}p_{\mathcal{I}}(m,v)$ belongs to B_t'', expansion (8.3.6) always holds since $|m_i| \wedge |m_j| \gg 1/\sqrt{t}$. Taking $|m_j| = 1/\log t$ for instance, we have

$$f_{t,C_{j,j}}(1/\log t) = \alpha \log t - \alpha \log(2|C_{i,j}|) - \log\log \sqrt{t} + o(1)\,.$$

as t tends to infinity. Given the uniformity over B_t'' in (8.3.2), this gives the asymptotic minimum of $f_{t,C_{j,j}}$, and ultimately $I(B_t'')$. ∎

We can now apply Theorem 5.1. Denote by

$$\rho_t = \sqrt{2I(B_t'') - \log(2\pi)^{d/2}}$$

the radius of the ball $\Lambda_{I(B_t)}$. As we have seen after the proof of Lemma 8.3.6,

$$\rho_t \sim \sqrt{2\alpha \log t} \qquad \text{as } t \to \infty\,.$$

In view of Lemma 8.3.4, for $\mathcal{I}$ in $J(C)$ and m belonging to $M_{\mathcal{I}}$, define

$$q_{\mathcal{I},t}(m) = \sum_{k\in\mathcal{I}} \operatorname{sign}(m_k) Q(tm_k^2) e_k$$

and its projection onto $\Lambda_{I(B_t)}$ through the normal flow,

$$r_{\mathcal{I},t}(m) = \rho_t \frac{q_{\mathcal{I},t}(m)}{|q_{\mathcal{I},t}(m)|}\,.$$

We consider the dominating manifold of dimension $k = 1$,

$$\begin{aligned}\mathcal{D}_{B_t''} = \bigcup_{\mathcal{I}\in J(C)} \Big\{ r_{\mathcal{I},t}(m) \,:\, m \in \bigcup_{\mathcal{I}\in J(C)} M_{\mathcal{I}}\,;\, m \notin \Omega_t\,;\\ \max_{1\leqslant i\leqslant d} |m_i| \leqslant \frac{\sqrt{t}\log\log t}{(\log t)^{1/\alpha}}\,, \tau_{B_t''}\big(r_{\mathcal{I},t}(m)\big) \leqslant \log\log t \Big\}\,.\end{aligned}$$

It is fairly clear that $\mathcal{D}_{B_t''}$ is a dominating manifold for the set $B_t'' \bigcap \Gamma_{I(B_t'')+M\log\log t}$. We can take M large enough such that

$$\int_{\Gamma^{\mathrm{c}}_{R(t)+M\log\log t}} e^{-I(y)}\mathrm{d}y = o\Big(\frac{\log t}{t^\alpha}\Big) \qquad \text{as } t \to \infty\,.$$

When applying the formula given in Theorem 5.1, the integral over $\mathcal{D}_{B_t''}$ splits into two parts. The first part comes from the contribution

of points $r_{\mathcal{I},t}(m)$ with $\mathcal{I}$ belonging to $J(C) \setminus J_*(C)$. This part is exactly like the integral we dealt with in the proof of Theorem 8.2.10. It is of order $1/t^\alpha$. Thus, we concentrate on the second part, coming from points $r_{\mathcal{I},t}(m)$ with $\mathcal{I}$ in $J_*(C)$.

Let $\mathcal{I} = \{\, i, j\,\}$ be in $J_*(C)$ and $r_{\mathcal{I},t}(m)$ be a point of $\mathcal{D}_{B_t''}$. From (8.3.2) we infer

$$\begin{aligned}\tau_{B_t''}\big(r_{\mathcal{I},t}(m)\big) &= \frac{1}{2}\big(|q_{\mathcal{I},t}|^2 - \rho_t^2\big) \\ &= \alpha \log |m_i m_j| - \alpha \log \gamma - \frac{1}{2} \log \frac{\log(\sqrt{t}|m_i|)}{\log \sqrt{t}} \\ &\qquad - \frac{1}{2} \log \frac{\log(\sqrt{t}|m_j|)}{\log \sqrt{t}} + o(1)\end{aligned}$$

Consequently,

$$\begin{aligned}&\exp\Big(- \tau_{B_t''}\big(r_{\mathcal{I},t}(m)\big)\Big) \\ &\qquad = \frac{\gamma^\alpha}{|m_i m_j|^\alpha} \sqrt{\Big(1 + \frac{\log |m_i|}{\log \sqrt{t}}\Big)\Big(1 + \frac{\log |m_j|}{\log \sqrt{t}}\Big)}\big(1 + o(1)\big)\end{aligned}$$

uniformly over the part of $\mathcal{D}_{B_t''}$ corresponding to points $r_{\mathcal{I},t}(m)$ with m in $M_{\mathcal{I}}$. Notice that if $C_{j,j}$ is nonzero, the inequality

$$\tau_{B_t''}\big(r_{\mathcal{I},t}(m)\big) \leqslant M \log\log t$$

required in $\mathcal{D}_{B_t''}$ imposes $|m_j| \leqslant (\log t)^{2M/\alpha}$ for t large enough; this can be seen from the above expression of $\tau_{B_t''}\big(r_{\mathcal{I},t}(m)\big)$ and using the same lower bound argument as in the proof of Lemma 8.3.6 to handle the term in $\log\big(\log(\sqrt{t}|m_i|)/\log\sqrt{t}\big)$. Therefore, when applying Theorem 5.1 to a component coming from $M_{i,j}$ with $C_{j,j}$ nonzero, we need only to integrate for $|m_j| \leqslant (\log t)^{2M/\alpha}$.

Still in order to apply the formula in Theorem 5.1, on $\mathcal{D}_{B_t''}$,

$$\big|\mathrm{D}I\big(r_{\mathcal{I},t}(m)\big)\big| = |r_{\mathcal{I},t}(m)| \sim \rho_t \sim \sqrt{2\alpha \log t} \qquad \text{as } t \to \infty\,.$$

We can now calculate the matrix $G_{B_t''}\big(r_{\mathcal{I},t}(m)\big)$. In the range of m's that we are considering in $\mathcal{D}_{B_t''}$, Lemma 8.3.4 shows that $\partial B_t''$ behaves like a ruled surface made of $(d-2)$-dimensional flat subspaces in directions orthogonal to $V_{\mathcal{I}}$. Hence, the very same argument as in Lemma 8.2.8 shows that

$$G_{B_t''}\big(r_{\mathcal{I},t}(m)\big) \sim \frac{\mathrm{Id}_{\mathbb{R}^{d-2}}}{\sqrt{2\alpha \log t}} \qquad \text{as } t \to \infty\,,$$

uniformly over the part of $M_{\mathcal{I}}$ we are considering. We can then write explicitly the contribution of the part related to $M_{\mathcal{I}}$ in the formula given in Theorem 5.1. This contribution is

$$e^{-I(B_t'')}(2\pi)^{\frac{d-2}{2}}\int \frac{e^{-\tau_{B_t''}}}{|DI|^{d/2}(\det G_{B_t''})^{1/2}}\,\mathrm{d}\mathcal{M}_{r_{\mathcal{I},t}(M_{\mathcal{I}})}\,,$$

where we integrate over all points $r_{\mathcal{I},t}(m)$, for m in $M_{\mathcal{I}}$, with $|m_i|\vee|m_j| \leqslant \sqrt{t}\log\log t/(\log t)^{1/\alpha}$; and $|m_j| \leqslant (\log t)^{2M/\alpha}$ if $C_{j,j}$ is nonzero. As in Theorem 8.2.10, we can replace the Riemannian measure on $r_{\mathcal{I},t}(M_{\mathcal{I}})$ by that on $q_{\mathcal{I},t}(M_{\mathcal{I}})$. Its expression in the local parameterization given by m_j in Lemma 8.3.2 is

$$\bigg(\Big(\frac{\mathrm{d}}{\mathrm{d}m_j}Q(tm_i^2)\Big)^2+\Big(\frac{\mathrm{d}}{\mathrm{d}m_j}Q(tm_j^2)\Big)^2\bigg)^{1/2}\mathrm{d}m_j\,.$$

Since $Q'(x)\sim\sqrt{\alpha}/(2x\sqrt{\log x})$ as x tends to infinity, this expression is equivalent to

$$\bigg(\frac{\alpha}{m_i^2\log(tm_i^2)}\Big(\frac{dm_i}{dm_j}\Big)^2+\frac{\alpha}{m_j^2\log(tm_j^2)}\bigg)^{1/2}\mathrm{d}m_j\,.$$

Putting all the pieces together, the contribution to the integral in Theorem 5.1 coming from $M_{\mathcal{I}}$ is

$$\begin{aligned}
&\frac{\log\sqrt{t}}{t^\alpha}\frac{K_{s,\alpha}^2\alpha^\alpha 4\pi}{(2\pi)^{d/2}\gamma^\alpha}(2\pi)^{(d-2)/2}\times\\
&\quad\int\frac{\gamma^\alpha}{|m_im_j|^\alpha}\frac{\sqrt{\Big(1+\dfrac{\log|m_i|}{\log\sqrt{t}}\Big)\Big(1+\dfrac{\log|m_j|}{\log\sqrt{t}}\Big)}}{(2\alpha\log t)^{d/4}(2\alpha\log t)^{-(d-2)/4}}\times\\
&\qquad\bigg(\frac{\alpha}{m_i^2\log(tm_i^2)}\Big(\frac{\mathrm{d}m_i}{\mathrm{d}m_j}\Big)^2+\frac{\alpha}{m_j^2\log(tm_j^2)}\bigg)^{1/2}\mathrm{d}m_j\,,
\end{aligned}\qquad(8.3.7)$$

where we integrate over m_j such that

$$\frac{(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha}}{2|C_{i,j}|\sqrt{t}}\big(1+o(1)\big)\leqslant|m_j|\leqslant(\log t)^{2M/\alpha}$$

if $C_{j,j}$ is nonzero, and

$$\begin{aligned}\frac{(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha}}{2|C_{i,j}|\sqrt{t}}\big(1+o(1)\big)&\leqslant|m_j|\\
&\leqslant\frac{\sqrt{t}}{(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha}}\end{aligned}$$

if $C_{j,j}$ is null. The finale of the proof consists in showing how simple this integral is, at least asymptotically! Up to multiplying it by 2, we can restrict the range of integration to m_j positive.

We first integrate in the range

$$\frac{(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha}\big(1+o(1)\big)}{2|C_{i,j}|} \leqslant m_j \leqslant \frac{1}{\log t}\,.$$

In this range, $m_i \sim 1/(2C_{i,j}m_j)$, and

$$\frac{1}{m_i^2}\Big(\frac{\mathrm{d}m_i}{\mathrm{d}m_j}\Big)^2 \sim \frac{1}{m_j^2}\,.$$

Using the change of variable $m_j = t^s$, the integral becomes — on that range —

$$\begin{aligned}
&2\frac{\log\sqrt{t}}{t^\alpha}\frac{K_{s,\alpha}^2\alpha^\alpha 2}{\sqrt{2\alpha\log t}}\int |2C_{i,j}|^\alpha\big(1{+}o(1)\big)\sqrt{\big(1-2s+o(1)\big)(1+2s)}\times\\
&\qquad\qquad \sqrt{\alpha}\,\frac{1}{t^s\sqrt{\log t}}\Big(\frac{1}{1-2s+o(1)}+\frac{1}{1+2s}\Big)^{1/2}\log t\; t^s\mathrm{d}s\\
&= 4\frac{\log\sqrt{t}}{t^\alpha}\frac{K_{s,\alpha}^2\alpha^\alpha}{\sqrt{2}}|2C_{i,j}|^\alpha\int\sqrt{2+o(1)}\,\mathrm{d}s\,,
\end{aligned}$$

where we integrate for

$$-\frac{1}{2}-\frac{M\log\log t}{2\alpha\log t}\big(1+o(1)\big) \leqslant s \leqslant -\frac{\log\log t}{\log t}\,.$$

This gives a term

$$2\frac{\log\sqrt{t}}{t^\alpha}K_{s,\alpha}^2\alpha^\alpha|2C_{i,j}|^\alpha\,.$$

When $1/\log t \leqslant m_j \leqslant (\log t)^{2M/\alpha}$, then $\log|m_i|$ and $\log|m_j|$ are $O(\log\log t)$. This part of the integral contributes a term less than

$$\begin{aligned}
&\frac{\log\sqrt{t}}{t^\alpha}O(1)\int\frac{2^\alpha|C_{i,j}|^\alpha}{(1-m_j^2C_{j,j})^\alpha}\frac{(\log\log t)^{1/2}}{(\log t)^{1/2}}\frac{1}{\sqrt{\log t}}\times\\
&\qquad\qquad\qquad\qquad \Big(\frac{1}{m_i^2}\Big(\frac{\mathrm{d}m_i}{\mathrm{d}m_j}\Big)^2+\frac{1}{m_j^2}\Big)^{1/2}\mathrm{d}m_j\,.
\end{aligned}$$

Since $C_{j,j}$ is nonpositive, $1-m_j^2C_{j,j}\geqslant 1$. Therefore

$$\Big|\frac{1}{m_i}\frac{\mathrm{d}m_i}{\mathrm{d}m_j}\Big| = \Big|\frac{1}{m_j}\frac{1+m_j^2C_{j,j}}{1-m_j^2C_{j,j}}\Big| \leqslant \frac{1}{m_j}\,.$$

The contribution in the integral is of order at most

$$O\Big(\frac{\log\log t}{t^\alpha}\Big)\int_{1/\log t}^{(\log t)^{2M/\alpha}}\frac{\mathrm{d}m_j}{m_j}=O\Big(\frac{(\log\log t)^2}{t^\alpha}\Big)=o\Big(\frac{\log t}{t^\alpha}\Big)\,.$$

Consequently, if $C_{j,j}$ is not zero, (8.3.7) is equivalent to

$$2\frac{\log\sqrt{t}}{t^\alpha}K_{s,\alpha}^2\alpha^\alpha|2C_{i,j}|^\alpha\,,$$

as t tends to infinity.

If $C_{j,j}$ vanishes, we need to add a contribution for the part where

$$(\log t)^{2M/\alpha}\leqslant m_j\leqslant\frac{\sqrt{t}}{(\log t)^{M/(2\alpha)}(\log\log t)^{1/\alpha}}\,.$$

We argue as in the case $m_j\leqslant 1/\log t$ above — this amounts to exploit the symmetry between m_i and m_j when m_j is large. Therefore, if $C_{j,j}$ is null, (8.3.7) is equivalent to

$$2\frac{\log\sqrt{t}}{t^\alpha}K_{s,\alpha}^2\alpha^\alpha|2C_{i,j}|^\alpha\,,$$

as t tends to infinity. Putting all the estimates together, we obtain

$$\int_{B_t''}e^{-I(y)}dy\sim 2\frac{\log\sqrt{t}}{t^\alpha}K_{s,\alpha}^2\alpha^\alpha\times\sum_{(i,j)\in J_*(C)}|2C_{i,j}|^\alpha\big(\mathrm{I}_{\mathbb{R}\setminus\{0\}}(C_{j,j})+2\mathrm{I}_{\{0\}}(C_{j,j})\big)\qquad(8.3.8)$$

Noticing that

$$\sum_{\{i,j\}\in J_*(C)}|2C_{i,j}|^\alpha\big(\mathrm{I}_{\mathbb{R}\setminus\{0\}}(C_{j,j})+2\mathrm{I}_{\{0\}}(C_{j,j})\big)=\sum_{i:C_{i,i}=0}\sum_{1\leqslant j\leqslant d}|2C_{i,j}|^\alpha\,,$$

we see that (8.3.8) is the expression given in the statement of Theorem 8.3.1 since we replaced C by $(C+C^{\mathrm{T}})/2$ in this proof.

To conclude the proof, we need to check the assumptions of Theorem 5.1. This part of the proof of Theorem 8.2.10 can be copied almost word for word, and this concludes the proof of Theorem 8.3.1. ∎

Combining Theorem 8.2.21 and Lemmas 8.3.2, we obtain the following result.

8.3.9. COROLLARY. *Let X be a d-dimensional random vector with independent and identically distributed components, all having a Student-like distribution with parameter α. Let C be a $d\times d$ matrix, with $N(C)>2$. Then*

$$\lim_{t\to\infty} t^\alpha P\{\,\langle CX,X\rangle \geqslant t\,\} = 0\,.$$

Proof. Notice that if the largest diagonal element of C is positive, then $N(C)=1$. Moreover, if this largest diagonal term vanishes, Lemma 8.3.2 gives $N(C)=2$. Then, $N(C)>2$ implies that all the diagonal coefficients of C are negative. Consequently, $N_0(C)$ — defined after (8.2.19) — is at least 2. Apply Theorem 8.2.21 to conclude. ∎

At this point, the reader who doubts of the usefulness of Theorem 5.1 in providing a systematic technique should try to obtain the result of this chapter by other methods. Maybe once the results are known, Theorems 8.2.10, 8.2.21 and 8.3.1 can be proved more simply. It is also hoped that though Theorem 8.2.1 can be derived by easier methods, the path taken makes the current proof rather didactic. As it may have been noticed, Theorems 8.2.10 and 8.3.1 add extra arguments to the basic ones developed to prove Theorem 8.2.1.

Notes

This chapter is motivated by the statistical applications in time series developed in chapter 11.

Concerning section 8.1, there is a classical argument for the Gaussian case. Replacing C by $(C+C^{\mathrm{T}})/2$, there is no loss of generality in assuming that C is symmetric. Thus we can diagonalize the matrix, writing $C=QDQ^{\mathrm{T}}$ for a diagonal matrix D and an orthogonal one Q. Define $Y=Q^{\mathrm{T}}X$. Since the standard Gaussian distribution is invariant under orthogonal transformation, Y has again a standard normal distribution. Thus $\langle CX,X\rangle = \sum_{1\leqslant i\leqslant d} Y_i^2 D_{i,i}$ is a weighted sum of independent chi-square random variables. This can be generalized to noncentered Gaussian distributions — see Imhoff (1961) — and opens the door for saddlepoint approximations — Barndorff-Nielsen (1990).

The orthogonal invariance argument can be used for other ad hoc distributions. However, the Gaussian one is the only orthogonally invariant distribution with independent marginals.

For quadratic forms with heavy tail distribution, not much seems to be known. Davis and Resnick (1986) contains some asymptotic results as d tends to infinity in a time series context, based on the point process technique exposed in chapters 3–4 of Resnick (1987).

This chapter 8 is full of open questions. To state a few, what happens in the degenerate cases when $N(C) > 2$? Can one find a closed formula for the integral involved in Theorem 8.2.10? What is a good numerical scheme to compute such integral? Can we find more terms and obtain an asymptotic expansion? Can one obtain good upper bounds instead of asymptotic equivalents? Answers to these last two questions would be useful in the applications developed in chapter 11. Can one prove conjecture 8.1.1?

When dealing with heavy tails, there is a fashionable extension which consists in replacing any power function by itself times a slowly varying function. This is done mainly for linear functions of random variables. For quadratic functions, things turn to be much more complicated, and the classical guess, consisting of putting the same slowly varying function in the tail equivalent, is plain wrong. This can be seen already when multiplying two heavy tail random variables.

9. Random linear forms

In this chapter, we investigate the following problem. A d-dimensional random vector $X = (X_1, \ldots, X_d)$ defines a random linear form $X(p) = \langle X, p \rangle$ on $\mathbb{R}^d$. Given a subset M of $\mathbb{R}^d$ — eventually M could be of dimension much smaller than d — we can look at the restriction of the linear form on M, and at the distribution of its supremum,

$$X(M) = \sup\{ \langle X, p \rangle : p \in M \} .$$

Writing

$$A_t = \Big\{ x \in \mathbb{R}^d : \sup_{p \in M} \langle x, p \rangle \geqslant t \Big\} = tA_1 ,$$

we see that

$$P\{ X(M) \geqslant t \} = \int_{tA_1} \mathrm{d}P$$

where P is the probability measure of X. Provided 0 is not in the closure of A_1, the sets $tA_1 = A_t$ are moving to infinity with t. Theorem 5.1 may provide tail approximations of the distribution of $X(M)$.

There has been a tremendous amount of work on this problem, but from a somewhat different perspective. Traditionally, the set M is parametrized as $M = \big\{ p(s) : s \in S \big\}$ for some set S and some function $p : S \subset \mathbb{R}^k \to \mathbb{R}^d$. When $k = 1$, the random variable $X(M)$ is the supremum of a linear stochastic process; for $k > 1$, it is the supremum of a linear random field. Though it is widely used, parameterizing the set M has the disadvantage of hiding the fact that $X(M)$ is completely parameterization free. Any surjective — even not smooth, not injective, not measurable, etc. — change of parameterization of M leaves $X(M)$ invariant.

Often, one is interested in the supremum norm

$$|X|(M) = \sup \big\{ |\langle X, p \rangle| : p \in M \big\} ,$$

more than in $X(M)$. Up to changing M into $M \cup (-M)$, it is enough to consider $X(M)$.

The key point to understand is that obtaining an approximation of $P\big\{ X(M) \geqslant t \big\}$ for large t covers several distinct questions. First,

notice that A_1 is the complement in $\mathbb{R}^d$ of the convex set

$$C_M = \{\, x \in \mathbb{R}^d : x(M) \leqslant 1 \,\} = \bigcap_{p \in M} \{\, x : \langle x, p \rangle \leqslant 1 \,\}$$

— an intersection of half spaces. So, a first question is to understand how the probability that $X(M)$ is large is related to the geometry of ∂C_M. A second question is to study how the geometry of ∂C_M is related to that of M. When dealing with a parameterization of M, a third question is how to read in the parameterization the geometric information we need about M.

Among some amusing features of our point of view, we cannot resist mentionning the degenerate case $M = \{\, (1, \ldots, 1) \,\} \subset \mathbb{R}^d$. Then, one has $X(M) = X_1 + \cdots + X_d$. From our point of view, approximating the tail distribution of a sum of random variables, approximating the tail distribution of the supremum of a linear process, approximating the distribution of the supremum of a linear random field are all the same problem.

As in the previous chapters, we will investigate light and heavy tail distributions. Our purpose is again to illustrate the use of Theorem 5.1 and to show that very different geometric features govern the tail behavior of $X(M)$ according to the distribution of X.

9.1. Some results on convex sets.

In this section we relate the set $M \subset \mathbb{R}^d$ to the convex set C_M. Notice that $C_{M \cup M/2} = C_M$. Hence, C_M does not characterize M. To study which part of M is characterized by C_M, we first prove that we can also assume M is closed.

9.1.1. LEMMA. *For any subset M of $\mathbb{R}^d$, the equality $C_{\mathrm{cl}M} = C_M$ holds. Moreover, C_M contains the origin if and only if M is unbounded.*

Proof. The inclusion of M in its closure implies that of $C_{\mathrm{cl}M}$ in C_M. On the other hand, any point p in the closure of M is the limit of a sequence of points p_n belonging to M. Any point x in C_M satisfies $\langle x, p_n \rangle \leqslant 1$ and consequently $\langle x, p \rangle \leqslant 1$. This proves the inclusion of C_M in $C_{\mathrm{cl}M}$.

The second statement in the lemma follows from the equivalence between the inclusion of the ball of radius ϵ centered at the origin in C_M and that of M in the ball of radius $1/\epsilon$ around the origin. ∎

In what follows, we assume that M is closed and bounded, i.e.,

$$M \text{ is compact in } \mathbb{R}^d.$$

Another way to think of this assumption is that we can take M to be closed since $X(M) = X(\mathrm{cl}M)$. Moreover, $X(M)$ is infinite almost surely if and only if M is unbounded and X is nondegenerate. In short, the behavior of $X(M)$ is trivial if and only if M is unbounded. So, we may as well assume M to be compact.

Recall that for a convex set C, there is a dense set in its boundary containing points for which the tangent space to ∂C is well defined — see, e.g., Schneider (1993). Consequently, it makes sense to define

$$\begin{aligned} M_0 = \mathrm{cl}\big\{\, p \in \mathbb{R}^d : \text{there exists } x \in \partial C_M\,,\ \langle x, p\rangle = 1\,, \\ T_x\partial C_M \text{ exists, and } p \perp T_x \partial C \,\big\} \end{aligned}$$

The next proposition shows that M_0 is the smallest closed set in M necessary to describe ∂C_M.

9.1.2. PROPOSITION. *Assume M is compact.*
(i) The inclusion $M_0 \subset M$ holds.
(ii) Moreover, $C_M = C_{M_0}$.
(iii) If M_1 is closed in M and $C_{M_1} = C_M$, then $M_0 \subset M_1$.

Proof. We will use the following claim: If x is in ∂C_M and $T_x\partial C_M$ exists, then there exists p in M, orthogonal to $T_x\partial C_M$, such that $\langle x, p\rangle = 1$. Indeed, for such x, there exists p in M with $\langle x, p\rangle = 1$. Let u be a tangent vector to C_M at x. We can find a curve $x(\epsilon)$ in ∂C_M such that $x(\epsilon) = x + \epsilon u + o(\epsilon)$. Since $x(\epsilon)$ is in C_M, we must have $1 \geqslant \langle x(\epsilon), p\rangle = 1 + \epsilon\langle u, p\rangle + o(\epsilon)$ as ϵ tends to 0. Thus, $\langle u, p\rangle$ vanishes, and indeed, p is orthogonal to $T_x\partial C_M$.
(i) Let q be a point in $\mathbb{R}^d$, such that there exists x in ∂C_M with $\langle q, x\rangle = 1$, the tangent space $T_x\partial C_M$ exists, and q is normal to ∂C_M at x. By the above claim, there exists p in M and orthogonal to $T_x\partial C_M$ such that $\langle x, p\rangle = 1$. Since M is compact, 0 is in the interior of C_M and $T_x\partial C_M$ is of dimension $d-1$. Consequently, p and q must be collinear. They are equal since $\langle x, p\rangle = \langle x, q\rangle = 1$. Thus q belongs to M. Since M is closed, the inclusion $M_0 \subset M$ follows.
(ii) The inclusion of M_0 in M implies that of C_M in C_{M_0}. To obtain the reverse inclusion, convexity of C_{M_0} and C_M shows that we just

need to prove $\partial C_{M_0} \subset \partial C_M$. Assume that $\partial C_{M_0} \setminus \partial C_M$ contains a point x. Then

$$\langle x, p\rangle \leqslant 1 \text{ for all } p \in M_0\,. \tag{9.1.1}$$

Since C_M is included in C_{M_0}, such an x cannot belong to C_M. Moreover, since C_M contains the origin and x is not in C_M, there exists λ in $(0,1)$ such that $y = \lambda x$ belongs to ∂C_M. For any positive ϵ, there exists y_ϵ in ∂C_M such that $|y - y_\epsilon| \leqslant \epsilon$ and $T_{y_\epsilon}\partial C_M$ exists. Using our claim, there exists p_ϵ in M, orthogonal to $T_{y_\epsilon}\partial C_M$ and such that $\langle p_\epsilon, y_\epsilon\rangle = 1$, i.e., p_ϵ belongs to M_0. Since M_0 is closed and is in the compact set M, it is compact. As ϵ tends to 0, the points p_ϵ admit a cluster point p belonging to M_0, thus belonging to M. Then, as $\lim_{\epsilon\to 0} y_\epsilon = y$,

$$1 = \langle p, y\rangle = \lambda\langle p, x\rangle\,.$$

Consequently, $\langle p, x\rangle = 1/\lambda > 1$, which contradicts (9.1.1). It follows that $\partial C_{M_0} \subset \partial C_M$, and therefore $C_{M_0} \subset C_M$.
(iii) is clear from (i) and (ii), and shows the minimality of M_0. ∎

Proposition 9.1.2 motivates the following definition.

DEFINITION. *We say that a set M is reduced if $M = M_0$.*

If ∂C_M has a well defined tangent space at x, it admits a unit normal vector $N(x)$ pointing outward from C. The claim in the proof of Proposition 9.1.2 shows that there exists p in M, orthogonal to $T_x\partial C_M$, and such that $\langle p, x\rangle = 1$. Such a p is collinear to $N(x)$ and satisfies $|p|\langle x, N(x)\rangle = 1$. In conclusion,

$$M_0 = \mathrm{cl}\{\, N(x)/\langle x, N(x)\rangle : x \in \partial C_M\,,\ T_x\partial C_M \text{ exists}\,\}\,. \tag{9.1.2}$$

This set is called — traditionally assuming that ∂C_M is smooth — the polar reciprocal of ∂C_M — see, e.g., Schneider (1993).

Whenever ∂C_M is locally a C^2-manifold, the following lemma shows that M_0 is also locally a C^2-manifold. Moreover, the second fundamental form of M_0 is related to that of ∂C_M.

9.1.3. LEMMA. *Let $x : U \subset \mathbb{R}^{d-1} \mapsto \partial C_M$ be a local parameterization of ∂C_M. Then, $x_0 = N \circ x/\langle x, N \circ x\rangle$ defines a local parameterization of M_0. If $(b_{i,j})_{1\leqslant i,j\leqslant d-1}$ is the the matrix of the second fundamental form of ∂C_M at x in this parameterization, then $b_{0,i,j} = b_{i,j}/\big(\langle N(x)\,, x\rangle|x|\big)$ is the second fundamental form of M_0 in the corresponding parameterization x_0.*

Proof. We follow closely Hasani and Koutroufiotis (1985). Let $u = (u_1, \ldots, u_{d-1}) \in U$. Define $x_i = \partial x/\partial u_i$ and $N_i = \partial N \circ x/\partial u_i$. The function $x_0 = N \circ x/\langle x, N \circ x\rangle$ defines a local parameterization of M_0 according to (9.1.2). Define $f = \langle x, N \circ x\rangle$. Since $\langle x_i, N \circ x\rangle = 0$, we have

$$x_{0,i} = \frac{\partial x_0}{\partial u_i} = \frac{N_i}{f} - \frac{\langle x, N_i\rangle}{f^2} N\,.$$

Moreover, x is normal to M_0 at x_0 since

$$\langle x_{0,i}, x\rangle = \frac{1}{f}\langle N_i, x\rangle - \langle x, N_i\rangle \frac{1}{f} = 0\,.$$

for all $i = 1, \ldots, d-1$. Thus, the components of the second fundamental form of M_0 are

$$\begin{aligned} b_{0,i,j} &= -\Big\langle \frac{\partial x_0}{\partial u_i}, \frac{1}{|x|}\frac{\partial x}{\partial u_j}\Big\rangle = -\Big\langle \frac{N_i}{f} - \frac{\langle x, N_i\rangle}{f^2} N, \frac{x_j}{|x|}\Big\rangle = -\frac{1}{f|x|}\langle N_i, x_j\rangle \\ &= \frac{b_{i,j}}{f|x|}\,. \end{aligned}$$

∎

In Lemma 9.1.3, we derived properties of M_0 from knowledge on ∂C_M. We will also need to go the other way, that is obtain some information on the boundary of C_M from the knowledge of M or M_0.

Our next result shows that whenever C_M is bounded, an half line starting at the origin can cut M_0 in at most one point.

9.1.4. LEMMA. *Assume that C_M is compact in $\mathbb{R}^d$. If q is in M_0, then $q\mathbb{R}^+ \bigcap M_0 = \{\, q\,\}$.*

Proof. Consider a point q in M_0 and λ positive such that λq belongs to M_0 as well. Representation (9.1.2) of M_0 implies that $q = \lim_{n\to\infty} N(x_n)/\langle x_n, N(x_n)\rangle$ and $\lambda q = \lim_{n\to\infty} N(y_n)/\langle y_n, N(y_n)\rangle$ for points x_n, y_n in ∂C_M at which ∂C_M has a well defined tangent plane.

Both ∂C_M and S_{d-1} are compact. Up to extracting a subsequence, we can assume that x_n, y_n, $N(x_n)$ and $N(y_n)$ converge respectively to x, y, N_1, N_2. Then $q = N_1/\langle x, N_1\rangle$ and $\lambda q = N_2/\langle y, N_2\rangle$. Since q and λq are collinear, we have $N_1 = N_2$ or $N_1 = -N_2$. Since C_M is convex and contains the origin, $\langle x, N_1\rangle$ and $\langle x, N_2\rangle$ are nonnegative. The positivity of λ implies $N_1 = N_2$.

Next, C_M being convex, it lies on one side of its tangent spaces. Thus $\langle x_n - y_n, N(x_n)\rangle \leqslant 0$. Taking the limit as n tends to infinity, we

obtain $\langle x-y, N_1\rangle \leqslant 0$. Permuting x and y yields $\langle y-x, N_2\rangle \leqslant 0$. Since $N_1 = N_2$, the vector $x - y$ is orthogonal to N_1 and $\langle x, N_1\rangle = \langle y, N_2\rangle$. Thus, λ equals 1. ∎

As a consequence of Lemma 9.1.4, the next result asserts that whenever C_M is compact and M_0 is a manifold, the space $p\mathbb{R}$ is transverse to T_pM_0 for a typical point p in M_0.

9.1.5. LEMMA. *If M is compact in $\mathbb{R}^d$ and M_0 is a manifold, then $\{\, p \in M_0 : p\mathbb{R} \subset T_pM_0 \,\}$ is nowhere dense in M_0.*

Proof. Assume that the set under consideration is dense in an open set U of M_0. Since M_0 is smooth, $p\mathbb{R}$ is included in T_pM_0 for every p in U. Up to considering an open subset of U, we can assume that U does not contain the origin. Thus, $p/|p|$ is a unit vector field in the tangent bundle of M_0. Let γ be an integral curve of this field, with $\gamma(0)$ in U. It satisfies the equation $\gamma' = \gamma/|\gamma|$. Consequently, $\gamma = |\gamma|\gamma'$. Differentiating, we obtain $\gamma' = \gamma' + |\gamma|\gamma''$. Hence γ'' vanishes since γ does not. Consequently, γ' is constant and $\gamma(s) = \gamma(0) + s\frac{\gamma(0)}{|\gamma(0)|} = \gamma(0)\bigl(1 + \frac{s}{|\gamma(0)|}\bigr)$. Hence the curve γ is in the ray $\gamma(0)\mathbb{R}$, which contradicts Lemma 9.1.4. ∎

Making use of Lemmas 9.1.1–9.1.3 and Proposition 9.1.2, we can reduce M to M_0. Moreover, M_0 is smooth if ∂C_M is. In the following, we assume that

$$M \text{ is reduced and is a smooth } m\text{-dimensional submanifold of } \mathbb{R}^d. \tag{9.1.3}$$

As announced, our next task is to relate the differential geometric properties of ∂C_M to those of M. For this purpose, we build a local parameterization of ∂C_M starting from one for M. Lemma 9.1.5 asserts that under (9.1.3), if C_M is compact, then, at a typical point p of M_0, the direction $p\mathbb{R}$ is not contained in the tangent space T_pM_0. Actually, the proof of Lemma 9.1.5 shows slightly more; namely, that we cannot have $p\mathbb{R}$ included in T_pM_0 along a submanifold of M_0. We will assume more, even when C is not compact; namely that given p_0 in M,

$$p\mathbb{R} \text{ is transverse to } T_pM \text{ for all } p \text{ in a neighborhood of } p_0. \tag{9.1.4}$$

Consider a local parameterization

$$\underline{p} : (u_1, \ldots, u_m) \in U \subset \mathbb{R}^m \mapsto \underline{p}(u_1, \ldots, u_m) \in M$$

of M around p_0. We can assume without any loss of generality that U is an open neighborhood of the origin and that $\underline{p}(0) = p_0$. For what follows, it is convenient to extend $\underline{p}$ to a map defined on a neighborhood of the origin of $\mathbb{R}^{d-1}$. Thus, let V be a neighborhood of 0 in $\mathbb{R}^{d-1-m}$ and consider

$$p : u \in U \times V \subset \mathbb{R}^m \times \mathbb{R}^{d-1-m} \mapsto p(u) = \underline{p}(u_1, \ldots, u_m)\,.$$

Under (9.1.4), to each point p near p_0, we can associate a unit normal vector

$$\nu(p) \in (T_pM + p\mathbb{R}) \ominus T_pM\,;$$

that is, $\nu(p)$ is normal to T_pM in $T_pM + p\mathbb{R}$.

Let $X_{m+1}(p), \ldots, X_{d-1}(p)$ be an orthonormal moving frame in $(T_pM + p\mathbb{R})^\perp = \mathbb{R}^d \ominus (T_pM + p\mathbb{R}) \equiv \mathbb{R}^{d-m-1}$. For u in $U \times V$, let

$$X(u) = \frac{\nu \circ p(u)}{\big\langle \nu \circ p(u), p(u) \big\rangle} + \sum_{m+1 \leqslant j \leqslant d-1} \big(\phi_j \circ p(u) + u_j \big) X_j \circ p(u)$$

where the function ϕ_j's are C^2 and such that

$$(u_{m+1}, \ldots, u_{d-1}) \in V \mapsto I\big(X(u_1, \ldots, u_m, u_{m+1}, \ldots, u_{d-1}) \big)$$

is minimum at 0 — the existence of these functions and their smoothness comes from the implicit function theorem and the smoothness of I. By construction, $X(u)$ is normal to $T_{p(u)}M$ for all u in $U \times V$.

Our next lemma asserts that if M is positively curved, then X defines a local parameterization of ∂C_M. Combined with Lemma 9.1.3, it allows us to parameterize M or ∂C_M, whichever is the most convenient.

9.1.6. LEMMA. *Under (9.1.3)–(9.1.4), if the second fundamental form of M relative to the normal field X is definite positive at every point, then $X(u)$ defines a local parameterization of ∂C_M. Moreover, $p(u)$ is outward normal at ∂C_M at $X(u)$.*

Proof. Denote by $\Pi_M^X(p)$ the second fundamental form of M at p along the normal field X. Notice first that $X(u)$ is orthogonal to $T_{p(u)}M$. Also,

$$\langle p(u), X(u) \rangle = 1\,,$$

and p is orthogonal to $\operatorname{span}\{\, X_{m+1}, \ldots, X_{d-1}\,\}$. Consequently, if $q(s)$ is a curve on M, parametrized by arc length, such that $q(0) = p(u)$ and $q'(0) = t$, then, as s tends to 0,

$$\begin{aligned}\langle q(s), X(u)\rangle &= 1 + \frac{s^2}{2}\langle q''(0), X(u)\rangle + o(s^2)\\ &= 1 - \frac{s^2}{2}\Pi_M^X\bigl(p(u)\bigr)(t,t) + o(s^2)\,.\end{aligned}$$

Thus, $s \mapsto \langle q(s), X(u)\rangle$ is maximal and equals 1 at $s = 0$. This proves that $X(u)$ is in ∂C_M and that $X(u)$ indeed defines a local parameterization of ∂C_M.

To prove that $p(u)$ is normal to ∂C_M, notice first that

$$\frac{\partial}{\partial u_i}X(u) = X_i(u) \qquad \text{for } i = m+1, \ldots, d-1\,.$$

Furthermore, writing $\nu_i = \dfrac{\partial}{\partial u_i}\nu \circ p(u)$ and $p_i = \dfrac{\partial}{\partial u_i}p(u)$, for $i = 1, \ldots, m$,

$$\begin{aligned}X_i = \frac{\partial}{\partial u_i}X = \frac{\nu_i}{\langle \nu, p\rangle} - \frac{\langle p, \nu_i\rangle}{\langle p, \nu\rangle^2}\nu &+ \sum_{m+1\leqslant j\leqslant d-1} \mathrm{d}\phi_j(p)\cdot p_i\, X_j\\ &+ \sum_{m+1\leqslant j\leqslant d-1} (\phi_j\circ p + u_j)\mathrm{d}X_j(p)\cdot p_i\,.\end{aligned} \tag{9.1.5}$$

From the very definition of X_j, for $j = m+1, \ldots, d-1$, we infer

$$\langle X_j \circ p(u), p(u)\rangle = 0 \qquad \text{for } j = m+1, \ldots, d-1\,. \tag{9.1.6}$$

Differentiating this equality yields

$$\langle \mathrm{d}X_j(p)\cdot p_i, p\rangle + \langle X_j\circ p, p_i\rangle = 0\,, \quad i = 1, \ldots m\,, \quad j = m+1, \ldots, d-1\,.$$

Therefore, since X_j is orthogonal to T_pM, we have

$$\langle \mathrm{d}X_j(p)\cdot p_i, p\rangle = 0 \qquad \text{for } j = m+1, \ldots, d-1\,. \tag{9.1.7}$$

Combining (9.1.5)–(9.1.7) yields $\langle X_i, p\rangle = 0$ for $i = 1, \ldots, m$. Thus, $p(u)$ is normal to ∂C_M at $X(u)$.

Since C_M contains the origin and $\langle p, x\rangle = 1$ is positive, the vector $p(u)$ must be pointing outward ∂C_M. ∎

We can now explain how to compute the second fundamental form of ∂C_M at $X(u)$. It is determined by its value on the basis X_i of the

tangent space. Let us denote by $N = p/|p|$ the outward unit normal vector field to ∂C_M. The proof of Lemma 9.1.6 shows that

$$
\begin{aligned}
\langle \mathrm{d}N \cdot X_i, X_j \rangle &= \frac{1}{|p|} \langle p_i, X_j \rangle \\
&= \begin{cases} 0 \quad \text{if } i = m+1, \ldots, d-1 \text{ or } j = m+1, \ldots, d-1 \\ \dfrac{\langle \nu_j, p_i \rangle}{|p| \langle \nu, p \rangle} + \dfrac{1}{|p|} \sum_{m+1 \leqslant r \leqslant d-1} (\phi_r \circ p + u_r) \langle p_i, \dfrac{\partial}{\partial u_j} X_r \rangle \text{ otherwise} \end{cases}
\end{aligned}
$$

Consequently, if Π_M^Y denotes the second fundamental form of M relative to a unit normal vector field Y, we have

$$
\begin{aligned}
&-\langle \mathrm{d}N \cdot X_i, X_j \rangle \\
&\quad = \begin{cases} \dfrac{\Pi_M^\nu(p_i, p_j)}{|p| \langle \nu, p \rangle} + \dfrac{1}{|p|} \sum_{m+1 \leqslant r \leqslant d-1} (\phi_r \circ p + u_r) \Pi_M^{X_r}(p_i, p_j) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } i, j = 1, \ldots, m. \\ 0 \qquad \text{otherwise.} \end{cases}
\end{aligned}
\tag{9.1.8}
$$

This second fundamental form vanishes whenever i or j is not in $\{ 1, \ldots, m \}$, expressing the fact that C_M is a ruled surface with nontrivial generators if $m < d-1$. Along a generator of dimension $d-1-m$, the set C_M is flat and has vanishing curvature.

Now, consider a convex function I on $\mathbb{R}^d$. Since we are able to relate points and geometry of ∂C_M to points and geometry of M, we should be able to relate points in ∂C_M which minimize I to some specific points in M. For this purpose, to a function f defined on $\mathbb{R}^d$ we associate the functions

$$
\begin{aligned}
f^\bullet(x) &= \sup \{ f(y) : \langle x, y \rangle = 1 \} , \\
f_\bullet(x) &= \inf \{ f(y) : \langle x, y \rangle = 1 \} .
\end{aligned}
$$

The basic properties of these transforms will be of some use and are stated in the following proposition. Notice that the statement (iv) in the Proposition does not require any smoothness.

9.1.7. PROPOSITION. *Let I be a convex function on $\mathbb{R}^d$, such that $\lim_{|x|\to\infty} I(x) = \infty$. Then,*
(i) $I_\bullet$ is continuous on $\mathbb{R}^d \setminus \{ 0 \}$ with $\lim_{|x|\to 0} I_\bullet(x) = \infty$. Moreover, if I is minimal at 0, then $\lim_{|x|\to\infty} I_\bullet(x) = I(0)$;
(ii) if $I(C_M^{\mathrm{c}}) = I(\partial C_M)$, then

$$
I(C_M^{\mathrm{c}}) = \inf \{ I_\bullet(p) : p \in M \} = I_\bullet(M) = I_\bullet(M_0) ;
$$

(even if M is not reduced)
(iii) points in C_M^c minimizing I correspond naturally to points in M_0 minimizing I in the sense that

$$\bigcup_{\{x\in C_M^c\,:\,I(x)=I(C_M^c)\}} \{\, m\in M_0 : \langle m,x\rangle = 1\,\}$$
$$= \{\, m\in M_0 : I_\bullet(m) = I_\bullet(M_0)\,\}\,;$$

(iv) for any convex function f with its minimum at 0, the equality $(f_\bullet)^\bullet = f$ holds.

Proof. (i) If $\langle x,p\rangle = 1$, then $1\leqslant |x||p|$. Consequently,

$$I_\bullet(x) \geqslant \inf\Big\{\, I(p) : |p| \geqslant 1/|x|\,\Big\}\,,$$

and

$$\lim_{|x|\to 0} I_\bullet(x) = \lim_{|p|\to\infty} I(p) = \infty\,.$$

Moreover, if $\langle x,p\rangle = 1$, then $p = \dfrac{x}{|x|^2} + \mathrm{Proj}_{x^\perp} p$. Therefore,

$$I(0) = \inf\Big\{\, I(p) : |p|\in\mathbb{R}^d\,\Big\} \leqslant I_\bullet(x) \leqslant \inf\Big\{\, I\Big(\frac{x}{|x|^2}+q\Big) : q\perp x\,\Big\}$$
$$\leqslant I\Big(\frac{x}{|x|^2}\Big)\,,$$

The equality $\lim_{|x|\to\infty} I_\bullet(x) = I(0)$ follows.

Let us now prove that $I_\bullet$ is continuous. Let p be a nonzero vector in $\mathbb{R}^d$. Let ϵ be positive and less than $|p|$. Consider a point q at distance ϵ from p. Since I is continuous and blows up at infinity, there exists x in $\mathbb{R}^d$ such that $\langle x,p\rangle = 1$ and $I(x) = I_\bullet(p)$. Then,

$$|\langle x,q\rangle - 1| = |\langle x, q-p\rangle| \leqslant |x|\epsilon\,.$$

Hence, there exists y in $\mathbb{R}^d$ such that $\langle y,q\rangle = 1$ and $|x-y|\leqslant \epsilon|x|/|q|$ — take $y = x - \big(\langle x,q\rangle - 1\big)q/|q|^2$. Therefore,

$$I_\bullet(q) \leqslant I(y) \leqslant \sup\Big\{\, I(v) : |v-x| \leqslant \frac{|x|}{|q|}\epsilon\,\Big\}\,.$$

Since I is continuous and p is nonzero, it follows that

$$\limsup_{q\to p} I_\bullet(q) \leqslant I(x) = I_\bullet(p)\,.$$

Next, consider a sequence p_k in the ball $B(p,\epsilon)$ of radius ϵ centered at p, converging to p, and such that $\lim_{k\to\infty} I_\bullet(p_k) = \liminf_{q\to p} I_\bullet(q)$. Let x_k be such that $\langle p_k, x_k\rangle = 1$ and $I_\bullet(p_k) = I(x_k)$. The function I is bounded on the set $q/|q|^2$ for q in $B(p,\epsilon)$ provided ϵ is strictly less than $|p|$. Then, the sequence $I_\bullet(p_k) \leqslant I\bigl(p_k/|p_k|^2\bigr)$ is bounded, and so is the sequence $I(x_k)$. Therefore, x_k is in a compact set $I^{-1}([0,c])$ for some positive c, and admits a clustering point x. After taking a subsequence, we can assume that x_k converges to x as k tends to infinity. Since $\langle x_k, p_k\rangle = 1$, we have $\langle x, p\rangle = 1$ and thus $I_\bullet(p) \leqslant I(x)$. Moreover, continuity of I implies

$$I_\bullet(p) \leqslant I(x) = \lim_{k\to\infty} I(x_k) = \lim_{k\to\infty} I_\bullet(p_k) = \liminf_{q\to p} I_\bullet(q)\,.$$

Overall, $\lim_{q\to p} I_\bullet(q) = I_\bullet(p)$ and $I_\bullet$ is continuous on $\mathbb{R}^d \setminus \{\,0\,\}$.
(ii) Let ϵ be positive and p in M such that $I_\bullet(p) \leqslant I_\bullet(M) + \epsilon$. There exists x in $\mathbb{R}^d$ such that $\langle x, p\rangle = 1$ and $I_\bullet(p) = I(x)$. Since $\langle x, p\rangle = 1$, the point x is not in the interior of C_M. Since I is continuous,

$$I(C_M^{\mathrm{c}}) = I\bigl((\mathrm{int} C_M)^{\mathrm{c}}\bigr) \leqslant I(x) = I_\bullet(p) \leqslant I_\bullet(M) + \epsilon\,.$$

Since ϵ is arbitrary, we established the inequality $I(C_M^{\mathrm{c}}) \leqslant I_\bullet(M)$.

Next, we use the assumption $I(C_M^{\mathrm{c}}) = I(\partial C_M)$. Continuity of I ensures that there exists x in ∂C_M such that $I(x) = I(\partial C_M)$. Hence, there exists p in the closure of M with $\langle p, x\rangle = 1$ — otherwise, x would not be on the boundary of C_M. Therefore, $I_\bullet(p) \leqslant I(x)$, and since $I_\bullet$ is continuous,

$$I_\bullet(\mathrm{cl} M) = I_\bullet(M) \leqslant I_\bullet(p) \leqslant I(x) = I(C_M^{\mathrm{c}})\,.$$

This proves $I(C_M^{\mathrm{c}}) = I_\bullet(M)$. Since $C_M = C_{M_0}$, we also obtain $I_\bullet(M) = I_\bullet(M_0)$. This proves assertion (ii). Notice also that we proved $I_\bullet(p) = I(x)$.
(iii) Let x be in ∂C_M minimizing I over C_M^{c}. Since M_0 is closed, $\langle x, m\rangle = 1$ for some m in M_0. For this m, if $I_\bullet(m) < I(x)$, then there exists y such that $\langle m, y\rangle = 1$ and $I(y) < I(x)$. Given how we choose x, the point y is not in C_M^{c}. Since $\langle m, y\rangle = 1$ and M_0 is reduced, y is in ∂C_M, and x does not minimize I over C_M^{c}. Consequently, $I_\bullet(m) = I(x)$. Then, assertion (ii) implies that $I_\bullet(m) = I_\bullet(M_0)$. This proves that in statement (iii), the set in the left hand side is included in that in the right hand side.

To prove the reverse inclusion, let m in M_0 minimizing $I_\bullet$. Since $\lim_{|x|\to\infty} I(x) = \infty$ and I is continuous, $I_\bullet(m) = I(x)$ for some x such

that $\langle m, x\rangle = 1$. Then, $I(x) = I_\bullet(M_0)$ and x is not in the interior of C_M. Then, assertion (ii) implies $I(x) = I(C_M^{\mathrm{c}})$. Consequently, the set in the right hand side of statement (iii) is included in that in the left hand side.

(iv) If $\langle x, p\rangle = 1$, then $f_\bullet(p) \leqslant f(x)$. Thus, $(f_\bullet)^\bullet(x) \leqslant f(x)$. Now seeking a contradiction, let x be such that $f(x)$ is positive and assume that $(f_\bullet)^\bullet(x) < f(x)$. Let H be a supporting hyperplane at x of the level set $\{\, y : f(y) \leqslant f(x)\,\}$. We can write $H = \{\, e\,\}^\perp$ for some e in $\mathbb{R}^d$. If e is orthogonal to x, then x is in H. Since f is convex, $f(x) \leqslant f(x+h)$ for any h in H. Taking $h = -x$ leads $f(0) \geqslant f(x) > 0$, a contradiction. Thus, e is not orthogonal to x.

Notice that

$$\{\, p : \langle p, x\rangle = 1\,\} = \Big\{\, \frac{x}{|x|^2} + y : y \perp x\,\Big\}\,.$$

Moreover, if y is orthogonal to x,

$$\begin{aligned}\Big\{\, q : \Big\langle \frac{x}{|x|^2} + y, q\Big\rangle = 1\,\Big\}&\\ = \Big\{\, \alpha x + (1-\alpha)\frac{y}{|y|^2} + z &: \alpha \in \mathbb{R}\,,\ z \perp \operatorname{span}(x,y)\,\Big\}\,.\end{aligned}$$

Going back to the definition of $f_\bullet$ and $(f_\bullet)^\bullet$, the assumption $(f_\bullet)^\bullet(x) < f(x)$ can be rewritten: There exists ϵ positive such that whenever y is orthogonal to x, the inequality $f\big(\alpha x + (1-\alpha)y|y|^{-2} + z\big) \leqslant f(x) - \epsilon$ holds for some real α and z orthogonal to $\operatorname{span}(x,y)$.

Consider the following y. If e is collinear to x, let y be any nonzero vector in $\{\, e\,\}^\perp$. Otherwise, let $y = a\mathrm{Proj}_{x^\perp} e + bx$, where a, b are such that

$$\begin{aligned}0 &= \langle y - |y|^2 x, e\rangle\\ &= -a^2\, |\mathrm{Proj}_{x^\perp} e|^2 \langle x, e\rangle + a\, \langle \mathrm{Proj}_{x^\perp} e, e\rangle + (b^2|x|^2 + b)\langle x, e\rangle\,.\end{aligned}$$

This quadratic equation in a always has a solution for $|b|$ large enough as well as $|b|$ small enough, since $\langle x, e\rangle$ is nonzero. Moreover, for $|b|$ small enough but nonzero, the solution is not 0, and y is not null.

This choice of y ensures that e is in the space spanned by x and y. Consequently, the orthocomplement of $\operatorname{span}(x,y)$ is in H. Moreover, for any real α,

$$\alpha x + (1-\alpha)\frac{y}{|y|^2} = x + \frac{1-\alpha}{|y|^2}(y - |y|^2 x) \in x + H\,.$$

Consequently, for any real α and any z orthogonal to $\mathrm{span}(x,y)$, the point $\alpha x + (1-\alpha)y|y|^{-2} + z$ belongs to $x+H$. Since f is convex,

$$f\Big(\alpha x + (1-\alpha)\frac{y}{|y|^2} + z\Big) \geqslant f(x)$$

contradicting $(f_\bullet)^\bullet(x) < f(x)$. ∎

Given Lemma 9.1.1 and Proposition 9.1.2, we can replace M_0 by M in statement (iii) of Proposition 9.1.5.

The explicit calculation of $I_\bullet$ depends of course on I. Notice however that homogeneity is preserved, as indicated in the following result.

9.1.8. LEMMA. *If I is positively α-homogeneous, then $I_\bullet$ is $-\alpha$-homogeneous.*

Proof. The result is straightforward since

$$\begin{aligned} I_\bullet(tx) = \inf\big\{\, I(y) : \langle tx, y\rangle = 1 \,\big\} &= \inf\big\{\, I(y/t) : \langle x, y\rangle = 1 \,\big\} \\ &= t^{-\alpha} I_\bullet(x)\,. \end{aligned}$$

∎

Finally, we calculate $I_\bullet$ in an important case for applications. For r positive , let $|x|_r = (\sum_{1\leqslant i\leqslant d} |x_i|^r)^{1/r}$. If r is larger than 1 this is the ℓ_r-norm of x.

9.1.9. LEMMA. *If $I(x) = c\big(|x_1|^\alpha + \cdots + |x_d|^\alpha\big)$, then $I_\bullet(p) = c/|p|_\beta^\alpha$ where α and β are conjugate — i.e., $\alpha^{-1} + \beta^{-1} = 1$.*

Proof. Without any loss of generality, we can assume that $c=1$. If $\langle p, x\rangle = 1$, Hölder's inequality yields $1 \leqslant |x|_\alpha |p|_\beta$, and so $I(x) = |x|_\alpha^\alpha \geqslant 1/|p|_\beta^\alpha$.

On the other hand, for $x_i = \mathrm{sign}(p_i)|p_i|^{\frac{1}{\alpha-1}}/|p|_\beta^\beta$, we have $\langle x, p\rangle = 1$ and $I(x) = 1/|p|_\beta^\alpha$. ∎

In the situation described in Lemma 9.1.9, one sees that $I_\bullet(M)$ is related to $\sup\big\{\, |p|_\beta : p \in M \,\big\}$, that is to the radius of the smallest ball in ℓ_β which contains M.

9.2. Example with a light tail.

In this section, we consider a random vector $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$, having a log-concave density $\exp(-I)$. Our first result is elementary. It is inspired by its Gaussian analogue, where $I(x) = \frac{1}{2}|x|^2 + \log(2\pi)^{d/2}$. It shows that under a growth control on I, we can estimate the exponential decay of $P\{\, X(M) \geqslant t \,\}$ as t tends to infinity.

9.2.1. PROPOSITION. *If M is compact and*

$$\lim_{|v|\to 0} \limsup_{|x|\to\infty} \frac{I(x+v)}{I(x)} \leqslant 1\,, \tag{9.2.1}$$

then

$$\lim_{t\to\infty} \frac{1}{I_\bullet(M/t)} \log P\Big\{\, X(M) \geqslant t \,\Big\} = -1\,.$$

Proof. Recall that the events $\big\{\, X(M) \geqslant t \,\big\}$ and $\{\, X \not\in tC_M \,\}$ are equal. Let $a_t = I(tC_M^{\mathrm{c}})$. Proposition 2.2 yields

$$\lim_{t\to\infty} \frac{1}{a_t} \log \Big\{\, X(M) \geqslant t \,\Big\} = -1$$

if and only if

$$\lim_{\epsilon\to 0} \liminf_{t\to\infty} \frac{1}{a_t} \log |tC_M^{\mathrm{c}} \cap \Gamma_{(1+\epsilon)a_t}| = 0\,. \tag{9.2.2}$$

Let ϵ be positive. Since I is convex and C_M is a neighborhood of the origin the points minimizing I over $\mathbb{R}^d$ are included in tC_M for t large enough. Consequently, for t large enough, there exists x_t in $\partial(tC_M)$ such that $I(tC_M^{\mathrm{c}}) = I(x_t)$.

Let η be a positive number. Consider a set of orthogonal vectors $v_1, \ldots, v_d$ in a supporting hyperplane of ∂C_M at x_t, such that $\eta/2 \leqslant |v_i| \leqslant \eta$ for all $i = 1, \ldots, d-1$. The supporting hyperplane is the orthocomplement of a unit vector e, pointing outward from ∂C_M. We define $v_d = \eta e$. Since M is compact, the interior of C_M contains the origin. Therefore, x_t tends to infinity with t. Consequently, if η is small enough, (9.2.1) implies that for t large enough

$$I(x_t + v_i) \leqslant (1+\epsilon) I(x_t) = (1+\epsilon) a_t\,.$$

Since $\Gamma_{(1+\epsilon)a_t}$ is convex, the simplex with vertices x_t, $x_t + v_i$, $1 \leqslant i \leqslant d$, is in $tC_M^{\mathrm{c}} \bigcap \Gamma_{(1+\epsilon)a_t}$. Its volume does not depend on t; it bounds the volume of $tC_M^{\mathrm{c}} \bigcap \Gamma_{(1+\epsilon)a_t}$ below. Consequently, (9.2.2) holds.

It remains to prove that $I(tC_M^{\mathrm{c}}) = I_\bullet(M/t)$. This is clear since $tC_M = C_{M/t}$, and Proposition 9.1.7 holds. ∎

Condition (9.2.1) looks good, but is far from being the best that we can obtain. In particular, it does not cover the function $I(x) = \exp(|x|^\alpha)$ with $\alpha > 1$. The following will do, but assumes that I is differentiable.

9.2.2. PROPOSITION. *In Proposition 9.2.1, one can replace assumption (9.2.1) by*

$$\lim_{\delta\downarrow 0}\limsup_{x\to\infty}\frac{e^{-\delta I(x)}\sup_{|u|\leqslant 1}|\mathrm{D}I(x+ue^{-\delta I(x)})|}{I(x)} = 0\,.$$

Notice that since I blows up at infinity, $e^{-\delta I(x)}$ is very tiny for large x. In essence, when $d = 1$, the new condition asserts that $I'e^{-\delta I}/I$ tends to 0. When I is large,

$$I'e^{-\delta I}/I \leqslant I'e^{-\delta I/2} = 2(e^{-\delta I/2})'/\delta\,.$$

Therefore, the only possible limit for $I'e^{-\delta I}/I$ as its argument tends to infinity is 0 — but in this discussion, nothing guarantees that the limit exists. This does not show that the new condition holds for any convex function; but it suggests that those which do not satisfy this condition are rather pathological.

Proof of Proposition 9.2.2. We follow the proof of Proposition 9.2.1. All what we need to do is to specify how to pick the vectors v_i, $1 \leqslant i \leqslant d$. The new assumption implies that there exists a function $\delta(x)$ tending to 0 at infinity, such that

$$\lim_{x\to\infty}\frac{e^{-\delta(x)I(x)}\sup_{|x-y|\leqslant\exp(-\delta(x)I(x))}|\mathrm{D}I(y)|}{I(x)} = 0\,.$$

Let $\eta = \exp\bigl(-\delta(x_t)I(x_t)\bigr)$, and let ϵ be an arbitrary positive number. The previous limit shows that

$$\eta\sup_{|y-x_t|\leqslant\eta}|\mathrm{D}I(y)| \leqslant \epsilon I(x_t)$$

for t large enough. With this new η, take the v_i's exactly as in the proof of Proposition 9.2.1. Since $|v_i|$ is at most η, we have, for $i = 1, \ldots, d$,

$$I(x_t + v_i) \leqslant I(x_t) + |v_i| \sup_{y:|x-y|\leqslant|v_i|} |\mathrm{D}I(y)| \leqslant I(x_t)(1+\epsilon)\,.$$

Consequently, the simplex with vertices x_t, $X_t + v_i$, $1 \leqslant i \leqslant d$ lies in $tC_M \bigcap \Gamma_{(1+\epsilon)I(x_t)}$. Its volume is of order η^d. Thus, to check (9.2.2), it suffices to show that

$$\lim_{t\to\infty} \frac{\log \eta}{I(x_t)} = \lim_{t\to\infty} -\delta(x_t) = 0\,.$$

This is plain from the definition of $\delta(\cdot)$. ∎

Clearly, one can do many variations on the theme, and get different conditions for the conclusion of Proposition 9.2.1 to hold.

To obtain a sharper result than in Propositions 9.2.1 or 9.2.2, that is to estimate $P\{\, X(M) \geqslant t \,\}$ and not its logarithm, we need further assumptions. Many results could be obtained under various hypotheses. We will suppose that the random vector X has density ae^{-I}, where

$$I \text{ is convex, } \alpha\text{-positively homogeneous for some } \alpha > 1. \tag{9.2.3}$$

Under (9.2.3), Theorem 7.1 settles more or less the question of approximating

$$P(tC_M^{\mathrm{c}}) = a \int_{tC_M^{\mathrm{c}}} e^{-I(x)} \mathrm{d}x\,.$$

Of course, we need to verify the assumptions of Theorem 7.1. Since C_M is convex, assumption (7.5) is always satisfied, while (7.1) is guaranteed by the boundedness of M. Thus, only (7.3) and (7.4) are left to check.

It does not seem that we can work out a general theory much further. But let us show how we can obtain a tail equivalent for $P\{\, X(M) \geqslant t \,\}$ from Theorem 7.1.

Proposition 9.2.1 and Lemma 9.1.8 imply

$$\lim_{t\to\infty} \frac{1}{t^\alpha} \log P\{\, X(M) \geqslant t \,\} = -I_\bullet(M)$$

and so we are done as far as the exponential term is concerned. Reading the formula in Theorem 7.1, we have $I(A_1) = I_\bullet(M)$.

We then need to relate the differential geometric quantities involved in Theorem 7.1 to those of M. From section 9.1, we infer that we can reduce M to M_0. So, let us assume that

M_0 is a closed, connected, m-dimensional submanifold of $\mathbb{R}^d$.

Since we can replace M by M_0, we will drop the subscript and write M instead of M_0 until the end of this section. Consider a local parameterization $p(u)$ of M. We have seen in section 9.1 that it induces a parameterization $X(u)$ on ∂C_M. Assume that

$$\mathcal{D}_{\bullet,M} = \big\{\, p \in M : I_\bullet(p) = I_\bullet(M) \,\big\}$$
is a k-dimensional submanifold of M.

We can choose the parameterization $p(\cdot)$ such that

$$(u_1, \ldots, u_k) \in U' \subset \mathbb{R}^k \mapsto p(u_1, \ldots, u_k, 0, \ldots, 0) \in \mathcal{D}_{\bullet,M}$$

is a local parameterization of $\mathcal{D}_{\bullet,M}$. With the notation of section 9.1, it follows from Lemma 9.1.6 and Proposition 9.1.7 that

$$(u_1, \ldots, u_k) \in U' \mapsto X(u_1, \ldots, u_k, 0, \ldots, 0) \in \mathcal{D}_{C_M^{\mathrm{c}}}$$

is a parameterization of $\mathcal{D}_{C_M^{\mathrm{c}}}$. To compute the Riemannian measure $\mathcal{M}_{\mathcal{D}_{C_M^{\mathrm{c}}}}$, we first compute the first fundamental form of the surface $\mathcal{D}_{C_M^{\mathrm{c}}}$. We obtain

$$\mathrm{d}\mathcal{M}_{\mathcal{D}_{C_M^{\mathrm{c}}}}\big(X(u)\big) = \Big(\det\big(\langle X_i(u)\,, X_j(u)\rangle\big)_{1\leqslant i,j\leqslant k}\Big)^{1/2} \mathrm{d}u_1 \wedge \cdots \wedge \mathrm{d}u_k\,,$$

where $X_i(u)$ is given in (9.1.5). The first fundamental form of ∂C_M involves inner products of ν, ν_i, X_i, $\mathrm{d}X_j \cdot p_i$, which can be expressed in term of the third fundamental form of M — or analogously, in term of the connection form on its normal frame principal bundle. In general, such an expression is rather involved; we will see how it simplifies in some cases.

To compute the curvature term $G_{C_M^{\mathrm{c}}}$ in Theorem 7.1, the comment after the statement of Theorem 7.1 shows that it equals

$$G(x) = \Pi^\pi_{\Lambda_{I(C_M^{\mathrm{c}})},x} - \Pi^\pi_{\partial C_M,x}$$

for all x in $\mathcal{D}_{C_M^{\mathrm{c}}}$. Since $\overline{G}(x) = \Pi_{\Lambda_{I(C_M^{\mathrm{c}})},x} - \Pi_{\partial C_M,x}$ vanishes on directions tangent to $\mathcal{D}_{C_M^{\mathrm{c}}}$, one way to compute $\det G(x)$ is actually to

diagonalize $\overline{G}(x)$, and take the product of its positive eigenvalues — by construction, the eigenvalues of $\overline{G}(x)$ are nonnegative; they are null only on the eigensubspace $T_x\partial C_M^{\mathrm{c}} \ominus T_x\mathcal{D}_{C_M^{\mathrm{c}}}$. Ultimately, we need to calculate $\det\big(G(x)-\lambda\mathrm{Id}\big)$. For this purpose, consider an orthonormal basis $e_1,\ldots,e_{d-1}$ of $T_x\partial(C_M^{\mathrm{c}}) = T_x\Lambda_{I(C_M^{\mathrm{c}})}$.

Denote by $\mathcal{X}$ the matrix obtained in writing the vectors $X_1,\ldots,X_{d-1}$ in the basis $e_1,\ldots,e_{d-1}$, that is $\mathcal{X} = (\langle e_i, X_j\rangle)_{1\leqslant i,j\leqslant n}$. This matrix is nonsingular and

$$\begin{aligned}\det(\overline{G}(x)-\lambda\mathrm{Id}) = 0 &\iff \det(\mathcal{X}^{\mathrm{T}}\overline{G}(x)\mathcal{X} - \lambda\mathcal{X}^{\mathrm{T}}\mathcal{X}) = 0\\ &\iff \det\big((\mathcal{X}^{\mathrm{T}}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\overline{G}(x)\mathcal{X} - \lambda\mathrm{Id}\big) = 0\,.\end{aligned}$$

How do we compute $\mathcal{X}^{\mathrm{T}}\overline{G}\mathcal{X}$ and $(\mathcal{X}^{\mathrm{T}}\mathcal{X})^{-1}$?

Write $N_{\partial C_M}$ for the outward unit normal to ∂C_M. Since

$$\mathcal{X}^{\mathrm{T}}\overline{G}(x)\mathcal{X} = \left(\left\langle\Big(\frac{\mathrm{D}^2 I(x)}{|\mathrm{D}I(x)|} - \mathrm{d}N_{\partial C_M}(x)\Big)X_i\,,X_j\right\rangle\right)_{1\leqslant i,j\leqslant n-1},$$

we first compute $\langle \mathrm{D}^2 I(x)X_i, X_j\rangle$, $1\leqslant i,j\leqslant d-1$. The terms $\langle \mathrm{d}N_{\partial C_M}X_i, X_j\rangle$, $1\leqslant i,j\leqslant d-1$ can be computed with formula (9.1.8) and depend on the curvature of M via its second fundamental forms Π_M^{ν} and $\Pi_M^{X_r}$, $m+1\leqslant r\leqslant d-1$.

The first fundamental form $\mathcal{X}^{\mathrm{T}}\mathcal{X} = (\langle X_i, X_j\rangle)_{1\leqslant i,j\leqslant n-1}$ can be computed in the same way. One may notice however that for $i = 1,\ldots,m$ and $j = m+1,\ldots,d-1$,

$$\langle X_i, X_j\rangle = \frac{\langle \nu_i, X_j\rangle}{\langle \nu, p\rangle} + \mathrm{d}\phi_j(p)\cdot p_i + \sum_{m+1\leqslant r\leqslant d-1}\phi_r\langle \mathrm{d}X_r\cdot p_i\,, X_j\rangle\,.$$

Moreover, if $i,j = m+1,\ldots,d-1$, then

$$\langle X_i, X_j\rangle = \delta_{i,j}\,. \qquad \text{(Kronecker symbol)}$$

Finally, if $i,j = 1,2,\ldots,m$, the expression of $\langle X_i, X_j\rangle$ involves again the third fundamental form of M.

At this point, it does not seem possible to push the abstract calculation much further. The author hopes that it is clear that the tail behavior of $X(M)$ is governed by the differential geometry of M immersed in $\mathbb{R}^d$, a somewhat known fact. All the calculations can be implemented on a computer.

Some simplifications may occur in some specific cases. We now discuss some important ones.

Assume for instance that I is a radial function, namely that

$$I(x) = J\big(|x|^2\big) \text{ for a function } r \in [0,\infty) \mapsto J(\sqrt{r}) \text{ which is convex on } \mathbb{R}^+\text{, and increasing.} \tag{9.2.4}$$

Under (9.2.4), I is minimal if and only if $|x|$ is. Thus, the normalization $I\big(X(u)\big)$ minimum at $u_{m+1} = \cdots = u_{d-1} = 0$ forces $\phi_j = 0$. Moreover, at the minimum, i.e., on $\mathcal{D}_{C_M^c}$, the norm $|X(u)|$ is minimum, and therefore,

$$X \perp X_i\,, \qquad i = 1, 2, \ldots, d-1 \text{ on } \mathcal{D}_{C_M^c}\,.$$

Since ν_i is orthogonal to X_i for $i = 1, 2, \ldots, d-1$, the vectors X and ν are collinear on $\mathcal{D}_{C_M^c}$. Consequently, X, ν, p are collinear on $\mathcal{D}_{C_M^c}$, and all point outward $\Lambda_{I(C_M^c)}$.

Notice also that $|\nu| = 1$ forces $\langle \nu, \nu_i\rangle = \langle p\,, \nu_i\rangle = 0$.

Equation (9.1.5) becomes very simple then, namely

$$X_i = \frac{\nu_i}{\langle \nu, p\rangle} = \frac{\nu_i}{|p|} \qquad \text{on } \mathcal{D}_{C_M^c} \text{ and for } i = 1, 2, \ldots, m\,.$$

It follows that along $\mathcal{D}_{C_M^c}$, the first fundamental form of ∂C_M is given by

$$\langle X_i, X_j\rangle = \begin{cases} |p|^{-2}\langle \nu_i\,, \nu_j\rangle\,, & i,j = 1,2,\ldots,m. \\ |p|^{-1}\langle \nu_i\,, X_j\rangle\,, & i = 1,2,\ldots,m,\ j = m+1,\ldots,d-1. \\ \delta_{i,j}\,, & i,j = m+1,\ldots,d-1. \end{cases}$$

The second fundamental form of ∂C_M also undergoes some simplifications. Indeed equation (9.1.8) becomes

$$\langle \mathrm{d}N_{\partial C_M} X_i, X_j\rangle = \begin{cases} 0 & \text{if } i \vee j = m+1,\ldots,d-1. \\ -|p|^{-2}\langle p_i, \nu_j\rangle & \text{if } i,j = 1,2,\ldots,m. \end{cases} \tag{9.2.5}$$

— remember that the normalization $u_{m+1} = \ldots = u_{d-1} = 0$ at the minimum forces $\phi_j = 0$ on $\mathcal{D}_{C_M^c}$.

REMARK. Be careful when using (9.2.5). The X_j's, for $j = m+1,\ldots,d-1$ have unit norm. But for $j = 1,\ldots,m$, the norm of X_j may not be 1. If one wants an expression of the fundamental form in a basis of unit vectors, one should divide X_j by its norm in (9.2.5).

Nothing guarantees that the X_j's for $j = 1, \ldots, m$ are orthogonal to the X_j's for $j = m+1, \ldots, d-1$. Thus to use (9.2.5) in computations, one needs to use some form of orthogonalization technique to obtain the matrix $\mathrm{d}N_{\partial C_M}$ in an orthonormal basis.

REMARK. If we rescale M to λM, then C_M becomes $C_{\lambda M} = C_M/\lambda$. So, the second fundamental form of $C_{\lambda M}$ should be proportional to λ. Consequently, the expression on the right hand side of (9.2.5) is homogeneous in λ. With our choice, the parameterization of $C_{\lambda M}$ is $X_\lambda(u) = X(u)/\lambda$. Therefore $X_{\lambda,i} = X_i/\lambda$. For $i, j = 1, 2, \ldots, m$, the left hand side of (9.2.5) written for $C_{\lambda M}$ reads

$$\langle \mathrm{d}N_{\partial C_{\lambda M}} X_{\lambda,i}\, , X_{\lambda,j} \rangle = \Big\langle \mathrm{d}N_{\partial C_{\lambda M}} \frac{x_i}{\lambda}\, , \frac{x_j}{\lambda} \Big\rangle \, .$$

Thus, for C_{tM}, formula (9.2.5) is

$$\langle \mathrm{d}N_{\partial C_{\lambda M}} X_i, X_j \rangle = \lambda \langle p_i\, , N_j \rangle / |p|^2 \, .$$

This is indeed homogeneous in λ since X_i and X_j do not depend on λ.

Similarly to what happens for the second fundamental form of ∂C_M along $\mathcal{D}_{C_M^{\mathrm{c}}}$, the second fundamental form of the level sets $\Lambda_{I(C_M^{\mathrm{c}})}$ undergoes a great simplification when I is a radial function. Since the level sets are spheres, their second fundamental form at X is

$$\Pi_{\Lambda_{I(C_M^{\mathrm{c}})}}(X)(X_i, X_j) = |X|^{-1} \langle X_i, X_j \rangle = |p| \langle X_i, X_j \rangle \qquad \text{on } \mathcal{D}_{C_M^{\mathrm{c}}}$$

— recall that for a radial I, we proved that $X(u)$ is collinear to $p(u)$, and so $1 = \langle X(u), p(u) \rangle = |X(u)|\, |p(u)|$.

Further simplifications may occur by a good choice of the parameterizations. For example, it may happen that $\mathcal{D}_{C_M^{\mathrm{c}}}$ is is parametrized by $(u_1, \ldots, u_k, 0, \ldots, 0)$, where $k = \dim \mathcal{D}_{C_M^{\mathrm{c}}}$. In the case $\dim M = 1$, we can take X_3 to be collinear to the torsion vector of the curve M. This ensures that $\nu_i = \nu'$ is orthogonal to $X_4, \ldots, X_{d-1}$. The case $m = d-1$ is also rather specific. There is a vast number of possible specializations where more or less remarkable formulas can be obtained. However, it is not obvious that such extensive developments would bring more insight. They may be worthwhile for some specific applications and we will see some in chapters 10–12.

To conclude, we mention that when Theorem 7.5 applies, it shows that the distribution of X/t given $X(M)$ larger than t converges to

a distribution supported by $\mathcal{D}_{C_M^c} = \{ x \in \partial C_M : I(x) = I(C_M^c) \}$ as t tends to infinity. To a point x in $\mathcal{D}_{C_M^c}$ correspond points m in M_0 such that $\langle x, m \rangle = 1$. Proposition 9.1.7.iii shows that as x varies in $\mathcal{D}_{C_M^c}$, these m's vary among the points in M_0 minimizing $I_\bullet$. This implies that when M is closed, $\{ m \in M_0 : X(m) = X(M) \}$ given $X(M) \geqslant t$ is a random closed set whose distribution given $X(M) \geqslant t$ tends to be concentrated on points in M_0 that minimize $I_\bullet$. When $X(M)$ is achieved at a unique point $m(X)$ in M_0, Theorem 7.5 even gives the limiting conditional distribution of $\arg\max_{m \in M_0} X(m)$ given $X(M) \geqslant t$, as t tends to infinity. It is the image measure by m of the limiting conditional distribution of X/t given $X \in tC_M$; this latter limiting distribution is given in Theorem 7.5. In particular, if $I_\bullet$ is minimum at a unique point m_* of M_0, then $\arg\max_{p \in M} X(p)$ given $X(M) \geqslant t$ converges in probability to m_* as t tends to infinity..

9.3. Example with heavy tail distribution.

In this section we consider a random vector $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$ with independent coefficients, all having a Student-like distribution with parameter α. Thus, the distribution of X_i is absolutely continuous with respect to the Lebesgue measure, and satisfies

$$P\{ X_i \leqslant -x \} \sim P\{ X_i \geqslant x \} \sim K_{s,\alpha} \alpha^{(\alpha-1)/2} x^{-\alpha}$$

as x tends to infinity, for some constant $K_{s,\alpha}$.

The estimation of $P\{ X \not\in C \}$ for an arbitrary convex set C containing a neighborhood of the origin turns out to be amazingly simple. Recall that $(e_1, \ldots, e_d)$ denotes the canonical basis of $\mathbb{R}^d$. It is convenient to introduce the following terminology.

DEFINITION. *Let S be a set in $\mathbb{R}^d$, containing a neighborhood of the origin. A point of the form λe_i on ∂S is called an axial point of ∂S.*

As their name suggests, axial points of ∂C are points on ∂C which lie on a canonical axis. Notice that if C is not the whole space $\mathbb{R}^d$, then ∂C has at least one axial point. Moreover, ∂C has at most $2d$ axial points. These lower and upper bound can be achieved. For instance, the half space $\{ x \in \mathbb{R}^d : x_1 \leqslant 1 \}$ has a unique axial point, e_1; and the centered unit ball has $2d$ of them, namely plus or minus the vectors of the canonical basis.

9.3.1. THEOREM. *Let C be a convex neighborhood of 0 in $\mathbb{R}^d$. Assume that C is not the whole space $\mathbb{R}^d$. Let X be a random vector with independent and identically distributed components having a Student-like distribution S_α, with parameter α, and such that $S_\alpha(0) = 1/2$. Then*

$$P\{ X \not\in tC \} \sim t^{-\alpha} K_{s,\alpha} \alpha^{(\alpha-1)/2} \sum_a |a|^{-\alpha} \qquad \text{as } t \to \infty\,,$$

where the sum $\sum_a$ is taken over all the axial points of ∂C.

Proof. Notice first that if ∂C is not smooth at some of its axial points, say $a_1, \ldots, a_k$, then we can sandwich C between two smooth convex sets with axial points $(1-\epsilon)a_1, \ldots, (1-\epsilon)a_k$ and $(1+\epsilon)a_1, \ldots, (1+\epsilon)a_k$ respectively. For these approximating convex sets, the asymptotic formula in Theorem 9.3.1 can be proved assuming that ∂C is smooth. We then let ϵ tend to 0. So, there is no loss of generality in assuming that ∂C is smooth, which we do from now until the end of the proof.

The proof of Theorem 9.3.1 is then essentially the same as that of Theorems 8.2.1 and 8.2.10. It is hoped that the reader will be convinced that our unifying formalism is actually quite convenient, even though, one more time, some specific examples could be treated more easily with ad hoc methods.

Let s_α denote the density of a single X_i. To prove Theorem 9.3.1, we need to approximate the integral

$$P\{ X \not\in tC \} = \int_{\mathbb{R}^d \setminus tC} s_\alpha(x_1) \ldots s_\alpha(x_d)\, \mathrm{d}x_1 \ldots \mathrm{d}x_d$$

for large t. As in the proof of Theorem 8.2.1, extend the function $\Phi^{\leftarrow} \circ S_\alpha$ to $\mathbb{R}^d$ by considering it acting componentwise on each coordinate. Making the change of variable $Y = \Phi^{\leftarrow} \circ S_\alpha(X)$ leaves us to approximate

$$\int_{\Phi^{\leftarrow} \circ S_\alpha(tC^{\mathrm{c}})} \frac{e^{-|y|^2/2}}{(2\pi)^{d/2}}\, \mathrm{d}y\,.$$

This leads us to consider the convex function

$$I(y) = \frac{|y|^2}{2} - \log(2\pi)^{d/2}$$

and the sets

$$A_t = tC^{\mathrm{c}}\,, \qquad \text{and} \qquad B_t = \Phi^{\leftarrow} \circ S_\alpha(tC^{\mathrm{c}})\,.$$

Our first proposition hereafter evaluates $I(B_t)$ and locates the points of interests in B_t as far as minimizing the function I is concerned. To this aim, define

$$\gamma = \min\{\,|a| : a \text{ axial point of } \partial C\,\}\,.$$

9.3.2. PROPOSITION. *As t tends to infinity, we have*

$$\begin{aligned} I(B_t) = \alpha\log t - \frac{1}{2}\log\log t + \alpha\log\gamma \\ - \log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) + \log(2\pi)^{d/2} + o(1)\,. \end{aligned}$$

If a is an axial point of ∂C, then

$$\tau_{B_t}\big(\Phi^{\leftarrow}\circ S_\alpha(ta)\big) = \alpha\log\big(|a|/\gamma\big) + o(1) \qquad \text{as } t\to\infty\,.$$

For any positive number M_1, the set $\partial B_t \cap \Gamma_{I(B_t)+M_1\log\log t}$ lies in an $O(\sqrt{\log\log t})$-neighborhood of the points $\Phi^{\leftarrow}\circ S_\alpha(ta)$, where a is an axial point of ∂C. Consequently, there exists a positive M_2 such that this set lies in the image through $\Phi^{\leftarrow}\circ S_\alpha$ of a $(\log t)^{M_2}$-neighborhood of the axial points of tC.

Proof. An easy application of the expansion for $(\Phi^{\leftarrow}\circ S_\alpha)^2$ given in Lemma A.1.5 gives for any axial point a of ∂C,

$$\begin{aligned} I\big(\Phi^{\leftarrow}\circ S_\alpha(ta)\big) = \alpha\log t - \frac{1}{2}\log\log t + \alpha\log|a| - \log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) \\ + \log(2\pi)^{d/2} + o(1) \end{aligned}$$

as t tends to infinity.

Let M_1 be a positive real number. Consider a point u in $C^{\rm c}$ such that

$$I\big(\Phi^{\leftarrow}\circ S_\alpha(tu)\big) \leqslant \alpha\log t + M_1\log\log t\,. \tag{9.3.1}$$

Consider $\eta < 1-(1/\sqrt{2})$. For t large enough, the inequality

$$\begin{aligned} I\big(\Phi^{\leftarrow}\circ S_\alpha(tu)\big) &\geqslant \alpha(1-\eta)\sum_{1\leqslant i\leqslant d} \mathrm{I}_{[t^{-\eta},\infty)}\big(|u_i|\big)\log\big(t|u_i|\big) \\ &\geqslant \alpha(1-\eta)^2\sharp\{\,1\leqslant i\leqslant d : |u_i|\geqslant t^{-\eta}\,\}\log t \end{aligned}$$

implies that for u as considered,

$$\sharp\{\,1\leqslant i\leqslant d : |u_i|\geqslant t^{-\eta}\,\} \leqslant (1-\eta)^{-2}\big(1+o(1)\big) < 2\,.$$

However, since u is not in C, we must have

$$\sharp\big\{\, 1\leqslant i\leqslant d : |u_i|\geqslant t^{-\eta}\,\big\}\geqslant 1$$

for t large enough. Consequently, u has exactly one coordinate larger than $t^{-\eta}$. Thus, it must be in a $t^{-\eta}$-neighborhood of the canonical axes in $\mathbb{R}^d$. Such a point is of the form $u=\lambda a+v$ for some axial point a of ∂C, some $\lambda\geqslant 1-t^{-\eta}|a|^{-1}$ and v orthogonal to a with $|v|\leqslant t^{-\eta}$. Consequently,

$$I\big(\Phi^{\leftarrow}\circ S_\alpha(tu)\big)=\frac{1}{2}\Phi^{\leftarrow}\circ S_\alpha\big(t\lambda|a|\big)^2+\frac{1}{2}|\Phi^{\leftarrow}\circ S_\alpha(tv)|^2+\log(2\pi)^{d/2}\,.$$

This expression is asymptotically minimum for $v=0$ and $\lambda=1$ and $|a|$ minimum, i.e., $|a|=\gamma$. Combined with Lemma A.1.5, this gives the value of $I(B_t)$ up to $o(1)$ as t tends to infinity. Then, the value for $\tau_{B_t}\big(\Phi^{\leftarrow}\circ S_\alpha(ta)\big)$ follows from Lemma A.1.5 as well.

Still assuming (9.3.1), we must have

$$\begin{aligned}\alpha\log t+M_1\log\log t\geqslant{}&\frac{1}{2}\Phi^{\leftarrow}\circ S_\alpha\Big(t|a|\big(1+o(1)\big)\Big)^2\\&+\frac{1}{2}\max_{1\leqslant i\leqslant d}\Phi^{\leftarrow}\circ S_\alpha(tv_i)^2+\log(2\pi)^{d/2}\,.\end{aligned}$$

Hence, using Lemma A.1.5 to approximate $\Phi^{\leftarrow}\circ S_\alpha\Big(t|a|\big(1+o(1)\big)\Big)^2$,

$$\max_{1\leqslant i\leqslant d}\Phi^{\leftarrow}\circ S_\alpha(tv_i)^2\leqslant(M_1+1)\log\log t$$

for t large enough, that is $\Phi^{\leftarrow}\circ S_\alpha(tu)=O(\log\log t)^{1/2}$. Therefore, $\Phi^{\leftarrow}\circ S_\alpha(tv)$ is indeed in an $O(\log\log t)^{1/2}$-neighborhood of $\Phi^{\leftarrow}\circ S_\alpha(ta)$.

Given the expression for $\Phi^{\leftarrow}\circ S_\alpha$ in Lemma A.1.5, we also must have

$$\max_{1\leqslant i\leqslant d}\alpha\log\big(t|v_i|\big)\leqslant(M_1+1)\log\log t\,,$$

that is, $\max_{1\leqslant i\leqslant d}|tv_i|\leqslant(\log t)^{(M_1+1)/\alpha}$ for t large enough, which is the last statement of the proposition. ∎

We can now try to calculate the asymptotic equivalent given by Theorem 5.1. Given the proof of Proposition 9.3.2, it is natural to try the projection of the axial points of ∂B_t onto $\Lambda_{I(B_t)}$ as a dominating manifold. So, let ρ_t be the radius of the ball $\Lambda_{I(B_t)}$, and set

$$\mathcal{D}_{B_t}=\Big\{\,\rho_t\frac{\Phi^{\leftarrow}\circ S_\alpha(ta)}{|\Phi^{\leftarrow}\circ S_\alpha(ta)|}: a\text{ axial point of }\partial C\,\Big\}\,.$$

Notice that because we assumed $S_\alpha(0) = 1/2$, the equality $\Phi^{\leftarrow}\circ S_\alpha(se_i) = \Phi^{\leftarrow}\circ S_\alpha(s)e_i$ holds for any canonical vector e_i of $\mathbb{R}^d$ and any real number s. Equivalently, we have

$$\mathcal{D}_{B_t} = \Big\{\, \rho_t \frac{a}{|a|} : a \text{ an axial point of } \partial C \,\Big\}\,.$$

The dimension of $\mathcal{D}_{B_t}$ is $k = 0$.

From the values of $I(B_t)$ in Proposition 9.3.2, we infer that

$$\begin{aligned}\rho_t &= \big[2\alpha\log t - \log\log t + 2\alpha\log\gamma - 2\log(K_\alpha \alpha^{\alpha/2} 2\sqrt{\pi}) + o(1)\big]^{1/2} \\ &\sim \sqrt{2\alpha\log t} \qquad \text{as } t\to\infty\,.\end{aligned}$$

Putting all the pieces together, and assuming that we can verify its assumptions, the approximation formula in Theorem 5.1 yields,

$$\begin{aligned}P(A_t) &\sim e^{-I(B_t)}(2\pi)^{(d-1)/2} \sum_a \frac{\exp\Big(-\alpha\log\big(|a|/\gamma\big)\Big)}{\rho_t^{(d+1)/2}\det\big(G_{B_t}(\rho_t a/|a|)\big)^{1/2}} \\ &\sim \frac{\sqrt{\log t}}{t^\alpha} K_{s,\alpha} \frac{\alpha^{\alpha/2}\sqrt{2}}{(2\alpha\log t)^{(d+1)/4}} \sum_a \frac{1}{|a|^\alpha\big(\det G_{B_t}(\rho_t a/|a|)\big)^{1/2}}\,,\end{aligned}$$

as t tends to infinity, with the sum taken over all the axial points a of ∂C. So, it remains for us to calculate $\det G_{B_t}(\rho_t a/|a|)$ for all the axial points of ∂C, and check the assumptions of Theorem 5.1. Our next lemma does half of the task.

9.3.3. LEMMA. *For any axial point a of ∂C,*

$$G_{B_t}(\rho_t a/|a|) \sim \frac{\mathrm{Id}_{\mathbb{R}^{d-1}}}{\sqrt{2\alpha\log t}} \qquad \textit{as } t\to\infty\,.$$

Proof. Let a be an axial point of ∂C and let u be orthogonal to a. Since the origin is in the interior of the convex set C, the line $a\mathbb{R}$ intersects ∂C transversally. Therefore, we can parameterize ∂C around a by a ball in $\{\,a\,\}^\perp$, centered at 0. In other words, there exists a smooth function $h_a : \{\,a\,\}^\perp \to \mathbb{R}$ such that

$$p_a : u \in \{\,a\,\}^\perp \mapsto p_a(u) = \big(1 + h_a(u)\big)a + u \in \partial C$$

defines a parameterization of ∂C around a — i.e., for $|u|$ small enough, it is a parameterization. Since u is orthogonal to a and $\Phi^{\leftarrow}\circ S_\alpha$ acts componentwise,

$$\Phi^{\leftarrow}\circ S_\alpha\big(tp_a(u)\big) = \Phi^{\leftarrow}\circ S_\alpha\Big(t\big(1+h(u)\big)\Big)\frac{a}{|a|} + \Phi^{\leftarrow}\circ S_\alpha(tu)\,.$$

If $I\big(\Phi^{\leftarrow}\circ S_\alpha(tp_a)\big) \leqslant I(B_t) + M_1\sqrt{\log\log t}$ — the domain which will interest us after we choose c_{B_t} — Proposition 9.3.2 asserts that $|u| \leqslant (\log t)^{M_2}/t$ for some M_2. Since ∂C is smooth, we have $|h(u)| \leqslant M_3|u| \leqslant M_3(\log t)^{M_2}/t$ for some positive M_3 and in this range of u. Consequently

$$\Phi^{\leftarrow}\circ S_\alpha\Big(t\big(1+h(u)\big)a\Big) = \Phi^{\leftarrow}\circ S_\alpha(ta) + o(1) \qquad \text{as } t\to\infty\,;$$

and in the range $|u| \leqslant (\log t)^{M_2}/t$,

$$\Phi^{\leftarrow}\circ S_\alpha\big(tp_a(u)\big) = \Phi^{\leftarrow}\circ S_\alpha(ta) + \Phi^{\leftarrow}\circ S_\alpha(tu) + o(1) \qquad (9.3.2)$$

as t tends to infinity. Up to the term in $o(1)$, this last equation defines a plane orthogonal to $\Phi^{\leftarrow}\circ S_\alpha(ta)$. Following the proof of Lemma 8.2.8, it follows that

$$G_{B_t}\big(\Phi^{\leftarrow}\circ S_\alpha(ta)\big) \sim \frac{\mathrm{Id}_{\mathbb{R}^{d-1}}}{\rho_t} \sim \frac{\mathrm{Id}_{\mathbb{R}^{d-1}}}{\sqrt{2\alpha\log t}}\,. \qquad \blacksquare$$

Lemma 9.3.3 implies

$$\det G_{B_t}\big(\rho_t a/|a|\big)^{1/2} \sim (2\alpha\log t)^{-(d-1)/4}\,, \qquad \text{as } t\to\infty\,.$$

With the estimate of $P(A_t)$ obtained before the statement of Lemma 9.3.3, we obtain

$$P(A_t) \sim \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{t^\alpha}\sum_a |a|^{-\alpha} \qquad \text{as } t\to\infty\,. \qquad (9.3.3)$$

This is the asymptotic equivalent given in Theorem 9.3.1. Thus, it remains for us to check the assumptions of Theorem 5.1. Taking the risk of making the rest of the proof boring, we will do it in a systematic way, showing that this is a rather easy task.

Our candidate for c_{B_t} is $c_t = (d+1)\log\log t$. Indeed, combining Propositions 2.1 and 9.3.2, we obtain

$$\begin{aligned} L\big(I(B_t)+c_t\big) &\leqslant c_0 e^{-I(B_t)-c_t}\big(1+I(B_t)+c_t\big)^d \\ &= o(t^{-\alpha}) \qquad \text{as } t\to\infty\,. \end{aligned}$$

Given (9.3.3), this choice of c_t guarantees that (5.4) holds.

Given Proposition 9.3.2 and the way we constructed $\mathcal{D}_{B_t}$ — using a projection on the sphere of radius ρ_t — the set $\underline{B}_{t,M}$ lies on

$O(\log\log t)$-neighborhood of $\mathcal{D}_{B_t}$ on $\Lambda_{I(B_t)}$. Since the radius of injectivity of $\Lambda_{I(B_t)}$ — a sphere of radius $\rho_t \sim \sqrt{2\alpha\log t}$ — is $\pi\rho_t/2$, assumption (5.2) holds for t large enough.

Assumption (5.5) is almost plain. Equation (9.3.2) is the analogue of Lemma 8.2.2 or 8.2.12. It shows that the boundary ∂B_t near an axis point $\Phi^{\leftarrow}\circ S_\alpha(ta)$ is a plane orthogonal to a up to an $o(1)$-term. The very same argument as that used in the proof of Theorems 8.2.1 and 8.2.10 ensures that (5.5) is verified here.

(5.6) is clear.

(5.7) follows in the very same way as in the proof of Theorem 8.2.1. We still have $t_{0,M}(p) = O(\log\log t)$ while $K_{\max}(q,t_0) = O(\log t)^{-1}$ for $q\in\mathcal{D}_{B_t}$.

(5.8) is plain, for a sphere has positive Ricci curvature in $\mathbb{R}^d$.

(5.9), (5.10) and (5.11) follow exactly as in the proof of Theorem 8.2.1.

(5.12) and (5.13) are plain as well given the proof of Theorem 8.2.1 or 8.2.10, and this concludes the proof of Theorem 9.3.1. ∎

REMARK. Note that if we drop the assumption $S_\alpha(0) = 1/2$ in Theorem 9.3.1, that is S_α of median zero, axial points are not mapped anymore to axial points by $\Phi^{\leftarrow}\circ S_\alpha$. The argument we developed would still be valid though, since the asymptotic expansion for $\Phi^{\leftarrow}\circ S_\alpha$ shows that the component of $\Phi^{\leftarrow}\circ S_\alpha(te_i)$ on $\{e_i\}^\perp$ is asymptotically negligible compared to that on $e_i\mathbb{R}$. Thus, the conclusion of Theorem 9.3.1 still holds true without assuming $S_\alpha(0) = 1/2$.

From Theorem 9.3.1, we deduce the following limiting behavior of the conditional distribution of X given $X\notin tC$.

9.3.4. COROLLARY. *Let C be any convex neighborhood of* 0 *in $\mathbb{R}^d$, such that $C\neq\mathbb{R}^d$. If X is a random vector with independent and identically distributed components having a Student-like distribution with parameter α, then the distribution of X/t given $X\notin tC$ converges weakly* to*

$$\sum_a |a|^{-\alpha}P_a\Big/\sum_a |a|^{-\alpha}\,,$$

where the sums are taken over all axial points a of ∂C and P_a is a Pareto distribution concentrated on $a\mathbb{R}^+$, whose cumulative distribution function is given by

$$P_a\big\{\,\langle X, a/|a|\rangle \geqslant |a|+\lambda\,\big\} = \frac{|a|^\alpha}{\big(|a|+\lambda\big)^\alpha}\,,\qquad \lambda\geqslant 0\,.$$

Proof. Let λ be a positive number. Consider the convex set

$$D = \{\, x : \langle x, a/|a| \rangle \leqslant |a| + \lambda \,\} .$$

The set of all axial points of $\partial(D^{\mathrm{c}} \cap C^{\mathrm{c}})$ is just $\{\, a\frac{|a|+\lambda}{|a|} \,\}$, and a is an axial point of the convex set C. Consequently, applying Theorem 9.3.1 with the convex set C and $(D^{\mathrm{c}} \cap C^{\mathrm{c}})^{\mathrm{c}}$ yields

$$\begin{aligned} P\{\, X \notin tD | X \notin tC \,\} &= P\big\{\, X \in t(D^{\mathrm{c}} \cap C^{\mathrm{c}}) \,\big\} \big/ P\{\, tC^{\mathrm{c}} \,\} \\ &\sim \big(|a| + \lambda\big)^{-\alpha} \Big/ \sum_a |a|^{-\alpha} . \end{aligned}$$

This is the result, for the conditional distribution of X/t given $X \notin tD$ converges trivially to P_a. ■

Let us now go back to the study of processes of the form $\langle X, p \rangle$, for p in a set M of $\mathbb{R}^d$. Theorem 9.3.1 gives us the asymptotic behavior of $P\big\{\, X(M) \geqslant t \,\big\}$ as t tends to infinity. To obtain a more readable statement, we need to express the axial points of ∂C_M in term of M. This is done in the next result.

9.3.5. PROPOSITION. *The set of all axial points of ∂C_M coincide with the set of vectors $\epsilon a_{\epsilon,i} e_i$ where ϵ is in $\{\, -1, +1 \,\}$, and i is such that $\epsilon \langle p, e_i \rangle > 0$ for some p belonging to M, and*

$$1/a_{\epsilon,i} = \sup \{\, \epsilon \, \langle p, e_i \rangle : p \in M \,\} .$$

Proof. Let a be an axial point of ∂C_M. Necessarily $a = \epsilon |a| e_i$ for some ϵ in $\{\, -1, 1 \,\}$, and e_i a vector of the canonical basis of $\mathbb{R}^n$. Since a is in ∂C_M, we have $\langle a, p \rangle \leqslant 1$ for all p in M, and $\sup \{\, \langle a, p \rangle : p \in M \,\} = 1$. Thus, $\sup \{\, \epsilon |a| \langle e_i, p \rangle : p \in M \,\} = 1$ and the result follows. ■

If we have a parameterization $f(t) = \big(f_1(s), \ldots, f_d(s)\big)$ of M indexed by s in some set S, it is particularly easy to relate the behavior of $f(s)$ to the geometry of the set M captured in Proposition 9.3.5. This yields immediately the following result, where the reader will notice that the function f is completely arbitrary — no need for measurability, or any kind of regularity whatsoever!

9.3.6. THEOREM. *Let S be a set, and $f(s) = \big(f_1(s), \ldots, f_d(s)\big)$ be a bounded function defined on S. Let X be a random vector in $\mathbb{R}^d$, with independent and identically distributed components having a Student-like distribution with parameter α. Let*

$$X(S) = \sup_{s\in S} X_1 f_1(s) + \cdots + X_d f_d(s)\,.$$

Then,

$$P\big\{\, X(S) \geqslant t \,\big\} \sim K_{s,\alpha} \frac{\alpha^{\alpha-1/2}}{t^\alpha} \sum_{\substack{1\leqslant i\leqslant d\\ \epsilon\in\{-1,1\}}} c_{\epsilon,i}^\alpha \qquad \textit{as } t\to\infty\,,$$

where

$$c_{\epsilon,i} = \sup\big\{\, \big(\epsilon f_i(s)\big)_+ : s\in S \,\big\}\,, \qquad \epsilon\in\{\,-1,1\,\}\,,\ i=1,\ldots,d\,.$$

Another way to interpret $c_{\epsilon,i}$ is in looking at the projection of M on the i-th canonical axis $e_i\mathbb{R}$. The value $c_{\epsilon,i}$ is 0 if this projection is concentrated on the set $-\epsilon e_i\mathbb{R}$; otherwise $c_{\epsilon,i}$ is the coordinate of the largest point of this projection.

It is quite amusing to notice the following. Set $f(s) = (1,\ldots,1)$ for all s. Then $X(S) = X_1 + \cdots + X_d$. Theorem 9.3.6 implies that

$$P\{\, X_1 + \cdots + X_d \geqslant t \,\} \sim \frac{K_\alpha \alpha^{\frac{\alpha-1}{2}}}{\lambda^\alpha} d\,.$$

Using again Theorem 9.3.6 for $d=1$, we then infer

$$P\{\, X_1 + \cdots + X_d \geqslant t \,\} \sim d\big(1 - S_\alpha(t)\big) = dP\{\, X_1 \geqslant t \,\}\,.$$

This is a known asymptotic identity showing that the Student-like distributions are subexponential!

Notes

This chapter is connected with a huge literature. To proceed in order, I first cannot quite believe that the transform $I \mapsto I_\bullet$ and its use are new. It allows one to read the minimization of a convex function on the complement of a convex set over its polar reciprocal. However, I have not found it in the literature. Similarly to what is called the infimum convolution in convex analysis, it would be natural to call

$I_\bullet$ the infimum Radon transform of I. The relation $(I_\bullet)^\bullet = I$ in Proposition 9.1.7.iv is an inversion formula for this infimum Radon transform.

Lemma 9.1.3 is essentially contained in Hassanis and Koutroufiotis (1985). Though I refer to Schneider (1993) for convexity theory, the differential viewpoint in section 9.1 is closer to Bruce and Giblin (1992). The polar reciprocal is sometimes called the pedal surface. It would be desirable to connect further the global properties of M_0 and C_M.

Proposition 9.2.1 is a generalization of the finite dimensional version of Fernique (1970) and Landau and Shepp (1970). There are many proofs in the Gaussian setting, and the lectures by Ledoux (1996) are most illuminating. The current literature on related problems is connected with notions such as concentration of measure and a set of inequalities: isoperimetric, Sobolev logarithmic, Poincaré. A couple of pointers to this literature are Talagrand (1995) and Bobkov and Ledoux (1997, 2000). However, it is not quite clear that Proposition 9.2.1 can be recovered from the existing results in the literature. An open question is if the conclusion of Proposition 9.2.1 holds for any convex functions.

The result of Proposition 9.2.1 also makes sense in infinite dimension, using of course the dual space to define $I_\bullet$. But the proof given here breaks down in infinite dimensions.

The part of section 9.2 following Proposition 9.2.1 is connected with a flourishing literature on the Gaussian case and a few related distributions such as the chi-square. The point of view given here is close to an abstraction of Diebolt and Posse (1996). A radically different line of investigation is in Piterbarg (1996). A most interesting survey of the literature on supremum of Gaussian processes is in Adler (2000).

In the heavy tail case, Theorem 9.3.6 seems to be part of the folklore; but I have not found it in the literature. Both Theorems 9.3.1 and 9.3.6 can be proved directly by ad hoc methods.

Concerning sums of heavy tailed random variables, Bingham, Goldie and Teugels (1987) contains invaluable material. I learned about those things in part in Broniatowski and Fuchs (1995). Much nicer results than the one presented in this section exist, including second and higher order formulas. But, unfortunately, the accuracy of these expansions is incredibly poor, especially when the tail parameter α is large.

It is interesting to rework the proofs of this specific chapter assuming that the cumulative distribution function of X_i is $1 - F(x) = x^{-\alpha}\ell(x)$ for some slowly varying function ℓ at infinity. Linearity of the functional considered here leads to neat results — but they depend heavily on the linearity!

10. Random matrices

In this chapter we consider a matrix $X = (X_{i,j})_{1\leqslant i,j\leqslant d}$ with random coefficients that are independent and have the same distribution. Many quantities associated to X are of interest. For instance, its trace $\mathrm{tr}(X)$, its determinant $\det(X)$, or its operator norm $\|X\|$. All these quantities have in general complicated distributions which cannot be calculated very explicitly. Hence, it makes sense to investigate their tail behavior.

Before going further, let us mention that the trace of X is nothing but a sum of independent and identically distributed random variables. Results from section 9 give the tail approximations for $P\big\{\,\mathrm{tr}(X) \geqslant t\,\big\}$ when the coefficients of M have a Weibull or a Student-like distributions. Therefore, in this section, we will concentrate on the determinant and on the norm of X. We will see that their tail behavior often turns to be quite interesting, if not fascinating.

Throughout this chapter, it will be convenient to think of matrices as vectors in $\mathbb{R}^{n^2}$ as well as linear operators acting on $\mathbb{R}^n$. In particular, we denote by $E^{i,j}$ the canonical orthonormal basis of $\mathbb{R}^{n^2}$ viewed as matrices. Thus, $E^{i,j}$ denotes the matrix with 1 on the (i,j)-entry, and 0 elsewhere. In other words.

$$E^{i,j} = (\delta_{(i,j),(k,l)})_{1\leqslant k,l\leqslant n}$$

where $\delta_{u,v}$ is the Kronecker symbol.

Let $\mathrm{M}(n,\mathbb{R})$ denote the set of all $n\times n$ matrices with real coefficients. Also, we write $\mathrm{GL}(n,\mathbb{R})$ for the group of all invertible matrices in $\mathrm{M}(n,\mathbb{R})$, that is the linear group.

Since our method is differential geometric, we will need the differential and Hessian of the determinant as a map from $\mathrm{GL}(n,\mathbb{R})$ to $\mathbb{R}$. For the sake of completeness, we recall them.

10.0.1. LEMMA. *For all x in $\mathrm{GL}(n,\mathbb{R})$ and h,k in $\mathrm{M}(n,\mathbb{R})$,*

$$\mathrm{D}\det(x)h = \det(x)\mathrm{tr}(x^{-1}h)\,,$$

and

$$\mathrm{D}^2\det(x)(h,k) = \det(x)\mathrm{tr}(x^{-1}h)\mathrm{tr}(x^{-1}k) - \det(x)\mathrm{tr}(x^{-1}hx^{-1}k)\,.$$

Proof. Let x be an invertible matrix. Since $\det(x+h) = \det(x)\det(\mathrm{Id}+x^{-1}h)$, it is enough to calculate $\mathrm{D}\det(\mathrm{Id})$ and $\mathrm{D}^2\det(\mathrm{Id})$.

Let $\mathfrak{S}_n$ denote the group of permutations of n elements. Define $\mathfrak{S}_n^0 = \{\,\mathrm{Id}\,\} \subset \mathfrak{S}_n$ and let $\mathfrak{S}_n^1$ be the subset of $\mathfrak{S}_n$ made of all transpositions. The signature of a permutation σ is $\epsilon(\sigma) = +1$ (resp. -1) if σ is the composition of an even (resp. odd) number of transpositions. We have

$$\det(\mathrm{Id}+sh) = \sum_{\sigma\in\mathfrak{S}_n} \epsilon(\sigma) \prod_{1\leqslant i\leqslant n} (\delta_{i,\sigma(i)} + sh_{i,\sigma(i)})\,.$$

This sum over $\mathfrak{S}_n$ can be decomposed as a sum over $\mathfrak{S}_n^0$, plus one over $\mathfrak{S}_n^1$, plus a remainder term. The sum over $\mathfrak{S}_n^0$ has a unique term,

$$\begin{aligned}
\epsilon(\mathrm{Id}) &\prod_{1\leqslant i\leqslant n} (\delta_{i,i} + sh_{i,i}) \\
&= 1 + s\sum_{1\leqslant i\leqslant n} h_{i,i} + s^2 \sum_{1\leqslant i<j\leqslant n} h_{i,i}h_{j,j} + O(s^3) \\
&= 1 + s\,\mathrm{tr}(h) + \frac{s^2}{2}\Bigl(\sum_{1\leqslant i,j\leqslant n} h_{i,i}h_{j,j} - \sum_{1\leqslant i\leqslant n} h_{i,i}^2\Bigr) + O(s^3) \\
&= 1 + s\,\mathrm{tr}(h) + \frac{s^2}{2}\Bigl(\mathrm{tr}(h)^2 - \sum_{1\leqslant i\leqslant n} h_{i,i}^2\Bigr) + O(s^3)\,.
\end{aligned}$$

Next, we also obtain

$$\begin{aligned}
\sum_{\sigma\in\mathfrak{S}_n^1} \epsilon(\sigma) &\prod_{1\leqslant i\leqslant n} (\delta_{i,\sigma(i)} + sh_{i,\sigma(i)}) \\
&= -\sum_{1\leqslant i<j\leqslant n} \prod_{\substack{1\leqslant k\leqslant n\\ k\notin\{i,j\}}} (1+sh_{k,k})s^2 h_{i,j}h_{j,i} \\
&= -\,s^2 \sum_{1\leqslant i<j\leqslant n} h_{i,j}h_{j,i} + O(s^3) \\
&= -\frac{s^2}{2}\Bigl(\sum_{1\leqslant i,j\leqslant n} h_{i,j}h_{j,i} - \sum_{1\leqslant i\leqslant n} h_{i,i}^2\Bigr) + O(s^3) \\
&= -\frac{s^2}{2}\Bigl(\mathrm{tr}(h^2) - \sum_{1\leqslant i\leqslant n} h_{i,i}^2\Bigr) + O(s^3)\,.
\end{aligned}$$

If σ is in $\mathfrak{S}_n \setminus (\mathfrak{S}_n^0 \cup \mathfrak{S}_n^1)$, at least 3 integers in $\{\,1,2,\ldots,n\,\}$ are not invariant under σ. For such permutation

$$\prod_{1\leqslant i\leqslant n} (\delta_{i,\sigma(i)} + \epsilon\delta_{i,\sigma(i)}) = O(s^3)\,.$$

It follows that

$$\det(\mathrm{Id} + sh) = 1 + s\,\mathrm{tr}(h) + \frac{s^2}{2}\big(\mathrm{tr}(h)^2 - \mathrm{tr}(h^2)\big) + O(s^3)$$

as s tends to 0. Consequently, for x in $\mathrm{GL}(n,\mathbb{R})$,

$$\mathrm{D}\det(x)(h) = \det(x)\mathrm{D}\det(\mathrm{Id})(x^{-1}h) = \det(x)\mathrm{tr}(x^{-1}h)\,.$$

Also, we have

$$\mathrm{D}^2\det(\mathrm{Id})(h,h) = \mathrm{tr}(h)^2 - \mathrm{tr}(h^2)$$

and by polarization

$$\mathrm{D}^2\det(\mathrm{Id})(h,k) = \mathrm{tr}(h)\mathrm{tr}(k) - \mathrm{tr}(hk)\,.$$

Consequently,

$$\begin{aligned}\mathrm{D}^2\det(x)(h,k) &= \det(x)\mathrm{D}^2\det(\mathrm{Id})(x^{-1}h, x^{-1}k)\\ &= \det(x)\big(\mathrm{tr}(x^{-1}h)\mathrm{tr}(x^{-1}k) - \mathrm{tr}(x^{-1}hx^{-1}k)\big)\end{aligned}$$

as claimed. ∎

We conclude this section by a trivial but useful formula. When needed, we write $\langle\cdot,\cdot\rangle_{\mathbb{R}^{n^2}}$ for the inner product in the Euclidean space $\mathbb{R}^{n^2}$. We use the tensor product notation, $E^{i,j}\otimes E^{l,m}$ to denote the bilinear form $x,y \in \mathrm{M}(n,\mathbb{R}) \mapsto E^{i,j}\otimes E^{l,m}(x,y) = \langle E^{l,m}, y\rangle_{\mathbb{R}^{n^2}}\langle E^{i,j}, x\rangle_{\mathbb{R}^{n^2}}$.

10.0.2. LEMMA. *On the basis $E^{i,j}$, the bilinear map $(h,k)\in \mathrm{M}(n,\mathbb{R}) \mapsto \mathrm{tr}(x^{-1}hx^{-1}k)$ has the form*

$$\sum_{1\leqslant i,j,l,m\leqslant n} (x^{-1})_{m,i}(x^{-1})_{j,l}E^{i,j}\otimes E^{l,m}\,.$$

Proof. It is straightforward,

$$\begin{aligned}\mathrm{tr}(x^{-1}hx^{-1}k) &= \sum_{1\leqslant i,j,l,n\leqslant n} (x^{-1})_{i,j}h_{j,l}(x^{-1})_{l,m}k_{m,i}\\ &= \sum_{1\leqslant i,j,l,m\leqslant n} (x^{-1})_{i,j}(x^{-1})_{l,m}E^{j,l}\otimes E^{m,i}(h,k)\,.\end{aligned}$$ ∎

Viewing $\mathrm{M}(n,\mathbb{R})$ as $\mathbb{R}^{n^2}$, we have a natural inner product

$$x, y \in \mathrm{M}(n,\mathbb{R}) \mapsto \langle x, y\rangle = \sum_{1\leqslant i,j\leqslant n} x_{i,j}y_{i,j} = \mathrm{tr}(x^{\mathrm{T}}y) = \mathrm{tr}(xy^{\mathrm{T}})\,.$$

Consequently, viewing $\mathrm{Ddet}(x)$ in $\mathbb{R}^{n^2}$, Lemma 10.0.1 implies that

$$\mathrm{Ddet}(x) = \det(x)(x^{-1})^{\mathrm{T}} \in \mathbb{R}^{n^2}\,. \tag{10.0.1}$$

10.1. Random determinants, light tails.

In this section, we consider a random matrix $X = (X_{i,j})_{1\leqslant i,j\leqslant n}$ where the $X_{i,j}$ are independent and identically distributed, each having a symmetric Weibull-like density

$$w_\alpha(u) = \frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)} \exp\Big(\frac{-|u|^\alpha}{\alpha}\Big)\,, \qquad u\in\mathbb{R}\,,\ \alpha>1\,.$$

One of the aim of this section is to show that the Gaussian case, obtained for $\alpha = 2$, is rather specific. The main reason is of course the invariance of the Gaussian distribution under the special orthogonal group.

Given the densities w_α of interest, let us define

$$I(x) = \frac{1}{\alpha}\sum_{1\leqslant i,j\leqslant n} |x_{i,j}|^\alpha\,, \qquad x\in \mathrm{M}(n,\mathbb{R}) \equiv \mathbb{R}^{n^2}\,.$$

Furthermore, define

$$A_t = \big\{\, x\in\mathrm{M}(n,\mathbb{R}) : \det X \geqslant t\,\big\} = t^{1/n}A_1\,.$$

Since I is homogeneous, we can use Theorem 7.1 in order to approximate

$$P\{\,\det X\geqslant t\,\} = \Big(\frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)}\Big)^{n^2}\int_{t^{1/n}A_1} e^{-I(x)}\,\mathrm{d}x\,.$$

This requires us to compute the dominating manifold $\mathcal{D}_{A_1}$. Unfortunately, I have not been able to do so in general. The following result will rely on an explicit calculation in some special cases, and a conjecture in general.

Our first lemma provides a necessary condition for a matrix to be in $\mathcal{D}_{A_1}$. We denote by $\mathrm{SL}(n,\mathbb{R})$ the special linear group on $\mathbb{R}^n$, that is the group of all matrices of determinant 1.

10.1.1. LEMMA. *If x is an $n\times n$ real matrix minimizing $I(x)$ subject to the constraint $\mathrm{det} x \geqslant 1$, then x is of determinant 1, that is belong to $\mathrm{SL}(n,\mathbb{R})$. Moreover, for such a matrix,*

$$(x^{-1})_{i,j} = \frac{1}{\lambda}\mathrm{sign}(x_{j,i})|x_{j,i}|^{\alpha-1}\,, \qquad 1\leqslant i,j\leqslant n\,, \tag{10.1.1}$$

where $n\lambda = \min\big\{\, I(x) : \det x = 1\,\big\}$.

Proof. Since $\det(\lambda x) = \lambda^n \mathrm{det} x$ and $I(\lambda x) = |\lambda|^\alpha I(x)$, we clearly have $\mathrm{det} x = 1$ at the constrained minimum. So, we need to find $\inf\big\{\, I(x) : \det x = 1\,\big\}$. At the minimum, the normal vector of the level set of I and $\{\, x : \det x = 1\,\}$ are collinear. Using (10.0.1), this condition writes

$$\Big(\frac{1}{\lambda}\mathrm{sign}(x_{i,j})|x_{i,j}|^{\alpha-1}\Big)_{1\leqslant i,j\leqslant n} = (x^{-1})^{\mathrm{T}}$$

for some nonzero λ.

Since $1 = (x^{-1}x)_{i,i}$ for all $i = 1,2,\ldots,n$, we obtain

$$\begin{aligned} n &= \sum_{1\leqslant i,k\leqslant n} (x^{-1})_{i,k} x_{k,i} = \frac{1}{\lambda}\sum_{1\leqslant i,k\leqslant n} \mathrm{sign}(x_{k,i})|x_{k,i}|^{\alpha-1}x_{k,i} \\ &= \frac{I(x)}{\lambda}\,, \end{aligned}$$

and the result follows. ∎

In general, I have been unable to solve (10.1.1) explicitely. But a very partial solution can be given, suggesting that the general one may be quite involved. As customary, we denote by $\mathrm{SO}(n,\mathbb{R})$ the special orthogonal group, that is the subgroup of $\mathrm{SL}(n,\mathbb{R})$ of all matrices X such that $X^{\mathrm{T}}X = \mathrm{Id}$.

10.1.2. LEMMA. *(i) If $\alpha = 2$, then $\mathcal{D}_{A_1} = \mathrm{SO}(n,\mathbb{R})$.*
(ii) If $n = 2$ and $\alpha < 2$, then

$$\mathcal{D}_{A_1} = \left\{ \begin{pmatrix} 0 & -\epsilon \\ \epsilon & 0 \end{pmatrix}, \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon \end{pmatrix} : \epsilon\in\{\,-1,1\,\} \right\}.$$

(iii) If $n = 2$ and $\alpha > 2$, then

$$\mathcal{D}_{A_1} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} \epsilon_1 & -\epsilon_2 \\ \epsilon_2 & \epsilon_1 \end{pmatrix} : \epsilon_1, \epsilon_2 \in \{ -1, 1 \} \right\}.$$

Proof. (i) for $\alpha = 2$, equation (10.1.1) becomes

$$(x^{-1})_{i,j} = \frac{x_{j,i}}{\lambda} \qquad 1 \leqslant i, j \leqslant n\,.$$

Hence, $x^{-1} = x^{\mathrm{T}}/\lambda$ and $\mathrm{Id} = x^{-1}x = x^{\mathrm{T}}x/\lambda$. Moreover, $\det x = 1$ thanks to Lemma 10.1.1.

When n is odd, we deduce that $1 = \det \mathrm{Id} = \lambda^{-n}$. Consequently, $\lambda = 1$ and $x^{-1} = x^{\mathrm{T}}$. This proves $\mathcal{D}_{A_1} \subset \mathrm{SO}(n, \mathbb{R})$. Since I is invariant under the action of $\mathrm{SO}(n, \mathbb{R})$, we have $\mathcal{D}_{A_1} = \mathrm{SO}(n, \mathbb{R})$ for n odd.

When n is even, we can also have $\lambda = -1$. But this implies $-\mathrm{Id} = x^{\mathrm{T}}x$. Since $x^{\mathrm{T}}x$ is symmetric nonnegative, this is impossible for real matrices.

(ii)–(iii): Set $x = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Using equation (10.1.1), we rewrite the equality $x^{-1}x = \mathrm{Id}$ as

$$\begin{aligned} \lambda &= \mathrm{sign}(a)|a|^{\alpha-1}a + \mathrm{sign}(b)|b|^{\alpha-1}b = |a|^{\alpha} + |b|^{\alpha} \\ 0 &= \mathrm{sign}(a)|a|^{\alpha-1}c + \mathrm{sign}(b)|b|^{\alpha-1}d \\ 0 &= \mathrm{sign}(c)|c|^{\alpha-1}a + \mathrm{sign}(d)|d|^{\alpha-1}b \\ \lambda &= \mathrm{sign}(c)|c|^{\alpha-1}c + \mathrm{sign}(d)|d|^{\alpha-1}d = |c|^{\alpha} + |d|^{\alpha}\,. \end{aligned} \tag{10.1.2}$$

Multiplying the second equality by ab, the third by cd and subtracting yields

$$\big(|a|^{\alpha} - |d|^{\alpha}\big)bc + \big(|b|^{\alpha} - |c|^{\alpha}\big)ad = 0\,.$$

But $1 = \det x = ad - bc$ implies then

$$\big(|a|^{\alpha} - |d|^{\alpha} + |b|^{\alpha} - |c|^{\alpha}\big)bc + |b|^{\alpha} - |c|^{\alpha} = 0\,.$$

At this stage, the first and last equations in (10.1.2) yields $|b|^{\alpha} = |c|^{\alpha}$, i.e., $|b| = |c|$, which then implies $|a| = |d|$. Set $c = \epsilon_1 b$ and $d = \epsilon_2 a$ for ϵ_1, ϵ_2 in $\{ -1, 1 \}$.

If bd is nonzero, the second equation in (10.1.2) gives, after multiplication by $|bd|$,

$$\mathrm{sign}(a)|d|^{\alpha}\epsilon_1 \mathrm{sign}(b)|b|^2 + \mathrm{sign}(b)|b|^{\alpha}\mathrm{sign}(d)|d|^2 = 0\,.$$

Thus, $\mathrm{sign}(a)\epsilon_1 = -\mathrm{sign}(d)$ and $|d| = |b| = |a| = |c|$. Hence, the matrix is of the form

$$x = |a| \begin{pmatrix} \epsilon_1 & \epsilon_3 \\ \epsilon_2 & \epsilon_4 \end{pmatrix} \qquad \text{with } \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 \in \{\,-1, 1\,\}\,.$$

The condition $1 = \det x = a^2(\epsilon_1\epsilon_4 - \epsilon_2\epsilon_3)$ forces $\epsilon_1\epsilon_4 - \epsilon_2\epsilon_3$ to be positive. Consequently, $\epsilon_1\epsilon_4 = +1$, and $1 = 2a^2$, i.e., $a = \pm 1/\sqrt{2}$. Thus,

$$x_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \epsilon_1 & -\epsilon_2 \\ \epsilon_2 & \epsilon_1 \end{pmatrix}$$

solves (10.1.2).

Next, if $bd = 0$, let us assume that, say, $d = 0$. Equation (10.1.2) reads

$$\begin{aligned} \lambda &= |a|^\alpha + |b|^\alpha \\ 0 &= \mathrm{sign}(a)|a|^{\alpha-1}c \\ \lambda &= |c|^\alpha\,. \end{aligned}$$

Since x is of determinant 1, the matrix x is not zero. Therefore, the relation $2\lambda = |a|^\alpha + |b|^\alpha + |c|^\alpha + |d|^\alpha > 0$ forces $c \neq 0$. Hence, $a = 0$ and $|b|^\alpha = \lambda$. The matrix is of the form

$$x_2 = |c| \begin{pmatrix} 0 & \epsilon_2 \\ \epsilon_1 & 0 \end{pmatrix}\,.$$

The condition $\det x_2 = 1$ forces $|c| = 1$ and $\epsilon_1 = -\epsilon_2 \in \{\,-1, 1\,\}$.

Finally, if $b = 0$, similar arguments yields a solution $x_3 = \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon \end{pmatrix}$, with ϵ in $\{\,-1, 1\,\}$.

We then have

$$\begin{aligned} \sum_{1 \leqslant i,j \leqslant 2} |(x_1)_{i,j}|^\alpha &= \frac{4}{2^{\alpha/2}} = 2^{2-\frac{\alpha}{2}}\,, \\ \sum_{1 \leqslant i,j \leqslant 2} |(x_k)_{i,j}|^\alpha &= 2\,, \qquad k = 2, 3\,. \end{aligned}$$

If $\alpha < 2$, then $2 < 2^{2-(\alpha/2)}$, while if $\alpha > 2$, we have the reverse inequality $2 > 2^{2-(\alpha/2)}$. The result follows. ∎

In general, I conjecture the following.

10.1.3. CONJECTURE. *If $\alpha \neq 2$, then $\inf\big\{\, I(x) : \det x = 1 \,\big\}$ is achieved at a finite number of matrices. Moreover, the difference of the*

two fundamental forms of $\Lambda_{I(\mathrm{SL}(n,\mathbb{R}))}$ *and* $\mathrm{SL}(n,\mathbb{R})$ *at those matrices is positive. Finally, if* $\alpha < 2$*, these matrices have unique nonzero elements on each row and each column, whose absolute value is* 1*; hence, up to signs, they are permutation matrices.*

Some numerical computations support Conjecture 10.1.3. Also, permutation matrices satisfy equation (10.1.1).

We can now state our approximation of the tail probability for $\det X$.

10.1.4. THEOREM. *Let* $X = (X_{i,j})_{1\leqslant i,j\leqslant n}$ *be a random matrix with independent and identically distributed coefficients, each having a symmetric Weibull like density* w_α*.*
(i) If $\alpha = 2$ *then, as* t *tends to infinity,*

$$P\{\det X \geqslant t\} \sim \frac{\pi^{(n-1)(n+2)/4}}{(2\pi)^{n^2/2}\sqrt{n}}\mathrm{Vol}\big(\mathrm{SO}(n,\mathbb{R})\big)e^{-nt^{2/n}/2}t^{(n^2-n-2)/2}\,.$$

(ii) If $\alpha > 2$*, under conjecture 10.1.3,*

$$P\{\det X \geqslant t\} \sim c_1 e^{-I(A_1)t^{\alpha/n}}t^{(\alpha(n^2-1)-2n^2)/2n} \qquad \textit{as } t\to\infty\,,$$

where $c_1 > 0$ *is a constant.*

REMARK. The constant c_1 in (ii) can be numerically computed as will be clear from the proof and Lemma 10.1.1. The volume of $\mathrm{SO}(n,\mathbb{R})$ in (i) is the volume when $\mathrm{SO}(n,\mathbb{R})$ is viewed as a submanifold of $\mathbb{R}^{n^2}$. This volume is the $n(n-1)/2$-dimensional Hausdorff-Lebesgue measure of the special orthogonal group.

When plugging $\alpha = 2$ in the exponent of t in (ii), we obtain $-1/n$, which is clearly different from the exponent of t in (i). Hence, the exponent of t has a discontinuity at $t = 2$.

When $\alpha = 2$ and $n = 1$, then $\det x = x$; the formula reads

$$P\{X \geqslant t\} \sim \frac{1}{\sqrt{2\pi}}e^{-t^2/2}t^{-1} \qquad \text{as } t\to\infty\,,$$

a well known fact! $\mathrm{SO}(1,R) = \{1\}$, and its volume measure is obtained by putting a Dirac mass at 1.

Proof of Theorem 10.1.4. The result is an application of Theorem 7.1.

Let us first determine the exponential term of the asymptotic equivalent. Since $A_t = t^{1/n} A_1$ and I is α-homogeneous, this term is $e^{-t^{\alpha/n} I(A_1)}$. The calculation of $I(A_1)$ relies upon Lemma 10.1.2 and Conjecture 10.1.3. For $\alpha = 2$, Lemma 10.1.2 implies that $I(A_1) = I(\mathrm{Id}) = n/2$. For $n = 2$, Lemma 10.1.2 gives $I(A_1) = 2/\alpha$ if $\alpha < 2$, while $I(A_1) = 4/(\alpha 2^{\alpha/2}) = 2^{2-\alpha/2}/\alpha$ if $\alpha > 2$.

In general, $I(A_1)$ can be computed numerically.

To obtain the polynomial term in t in the asymptotic expansion, Theorem 7.1 requires us to calculate $k = \dim \mathcal{D}_{A_1}$. When $\alpha = 2$, we obtain $k = \dim \mathrm{SO}(n, \mathbb{R}) = n(n-1)/2$. For $\alpha \neq 2$, we have $k = 0$ since $\mathcal{D}_{A_1}$ is discrete — here we use Conjecture 10.1.3 when $n > 2$.

It remains to evaluate the constant c_1 in Theorem 7.1 and to verify the assumptions of Theorem 7.1. Since the differential geometries of $\mathcal{D}_{A_1}$ and ∂A_1 are involved as well as that of $\Lambda_{I(A_1)}$, we need to calculate the differential and Hessian of det and I. Lemma 10.0.1 takes care of the former. When dealing with matrices it is convenient to express $\mathrm{D}I$ and $\mathrm{D}^2 I$ on the orthonormal basis $E^{i,j}$. Thinking of $E^{i,j}$ as an element in the dual of $\mathbb{R}^{n^2} \equiv \mathrm{M}(n, \mathbb{R})$, we have $E^{i,j}(M) = \langle E^{i,j}, M \rangle = M_{i,j}$ for any matrix $M = (M_{i,j})$ in $\mathrm{M}(n, \mathbb{R})$. With this notation, the gradient and Hessian of I have the following form.

10.1.5. LEMMA. *For any $\alpha \geqslant 1$ and any $n \times n$ real matrix x,*

$$\mathrm{D}I(x) = \textstyle\sum_{1 \leqslant i,j \leqslant n} \mathrm{sign}(x_{i,j}) |x_{i,j}|^{\alpha-1} E^{i,j} .$$

Moreover, if $\alpha \geqslant 2$,

$$\mathrm{D}^2 I(x) = (\alpha - 1) \textstyle\sum_{1 \leqslant i,j \leqslant n} |x_{i,j}|^{\alpha-2} E^{i,j} \otimes E^{i,j} \in \mathrm{M}(n^2, \mathbb{R}) .$$

Proof. Viewing $\mathrm{M}(n, \mathbb{R})$ as $\mathbb{R}^{n^2}$, we have for every x, h in $\mathrm{M}(n, \mathbb{R})$,

$$\begin{aligned} I(x + \epsilon h) = I(x) + \epsilon & \sum_{1 \leqslant i,j \leqslant n} \mathrm{sign}(x_{i,j}) |x_{i,j}|^{\alpha-1} h_{i,j} \\ & + \frac{\epsilon^2}{2} \sum_{1 \leqslant i,j \leqslant n} (\alpha - 1) |x_{i,j}|^{\alpha-2} h_{i,j}^2 + O(\epsilon^2) \end{aligned}$$

as ϵ tends to 0. Since $h_{i,j} = \langle E^{i,j}, h \rangle$, we obtain the expression for $\mathrm{D}I$. Using the polarization formula to express $\mathrm{D}^2 I(x)(h, k)$, we see that

$$(h_{i,j} + k_{i,j})^2 - h_{i,j}^2 - k_{i,j}^2 = 2 h_{i,j} k_{i,j} = 2 E^{i,j} \otimes E^{i,j}(h, k) .$$

This gives the expression for $\mathrm{D}^2 I(x)$. ∎

We are equipped to determine the tangent spaces to $\mathcal{D}_{A_1}$ and $\Lambda_{I(A_1)}$, from which we will deduce $\det G_{A_1}$.

10.1.6. LEMMA. *For any x in $\mathcal{D}_{A_1}$,*

$$\begin{aligned} T_x\Lambda_{I(A_1)} &= \{\, h \in \mathrm{M}(n,\mathbb{R}) : \langle x^{-1\mathrm{T}}, h\rangle = 0 \,\} \\ &= \{\, x^{-1\mathrm{T}} \,\}^\perp = \{\, xh : h \in \mathrm{M}(n,\mathbb{R})\,,\ \mathrm{tr}\, h = 0 \,\}\,; \\ T_x\mathcal{D}_{A_1} &= \emptyset \textit{ if } \alpha \neq 2 \textit{ (under conjecture 10.1.3 for } n \neq 2\textit{).};\\ T_x\mathcal{D}_{A_1} &= \{\, xh : h \textit{ skewsymmetric} \,\} \textit{ if } \alpha = 2\,. \end{aligned}$$

Consequently, at any x of $\mathcal{D}_{A_1} \cap \Lambda_{I(A_1)}$,

$$T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1} = \begin{cases} \{\, x^{-1\mathrm{T}} \,\}^\perp & \textit{if } \alpha \neq 2 \\ \{\, xh : h \in \mathrm{M}(n,\mathbb{R})\,,\ h \textit{ symmetric}\ ,\ \mathrm{tr}(h) = 0 \,\} & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if } \alpha = 2. & \end{cases}$$

Proof. Since $\Lambda_{I(A_1)}$ is a level set of I, we have for all x in $\mathcal{D}_{A_1}$,

$$T_x\Lambda_{I(A_1)} = \{\, \mathrm{D}I(x) \,\}^\perp = \{\, x^{-1\mathrm{T}} \,\}^\perp$$

thanks to Lemmas 10.1.5 and 10.1.1. Since

$$\begin{aligned} \{\, x^{-1\mathrm{T}} \,\}^\perp &= \{\, h : \langle x^{-1\mathrm{T}}, h \rangle = 0 \,\} = \{\, h : \mathrm{tr}(x^{-1}h) = 0 \,\} \\ &= \{\, xh : \mathrm{tr}(h) = 0 \,\} \end{aligned}$$

the expressions for $T_x\Lambda_{I(A_1)}$ follow.

When α is different than 2, Conjecture 10.1.3 asserts that $\mathcal{D}_{A_1}$ is a finite set, and indeed $T_x\mathcal{D}_{A_1} = \emptyset$.

For α equal to 2, the dominating manifold is $\mathrm{SO}(n,\mathbb{R})$. The Lie algebra of $\mathrm{SO}(n,\mathbb{R})$ is the set of all skewsymmetric matrices — see, e.g., Knapp, 1996, §I.1 — and the expression for $T_x\mathcal{D}_{A_1}$ follows in this case.

The result on $T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1}$ is then clear since the skewsymmetric matrices are orthogonal to the symmetric ones. ∎

In order to describe the matrix G_{A_1} involved in Theorem 7.1, recall that the ℓ_p-norm of a vector $x \in \mathbb{R}^{n^2} \equiv \mathrm{M}(n,\mathbb{R})$ is

$$|x|_p = \Big(\sum_{1 \leqslant i,j \leqslant n} |x_{i,j}|^p \Big)^{1/p}\,.$$

10.1.7. LEMMA. *For x in $\mathcal{D}_{A_1}$, the matrix $G_{A_1}(x)$, is obtained in restricting the bilinear form*

$$|x|_{2(\alpha-1)}^{1-\alpha}(\alpha-1)\sum_{1\leqslant i,j\leqslant n}|x_{i,j}|^{\alpha-2}E^{i,j}\otimes E^{i,j} + |x|_{2(\alpha-1)}^{1-\alpha}\lambda^{-1}\sum_{1\leqslant i,j,k,l\leqslant n}(x^{-1})_{l,i}(x^{-1})_{j,k}E^{i,j}\otimes E^{k,l}$$

to the subspace $T_x\Lambda_{I(A_1)}\ominus T_x\mathcal{D}_{A_1}$.

Proof. Given the comment following Theorem 7.1, it is enough to calculate the second fundamental form of the hypersurface $\Lambda_{I(A_1)}$ (resp. ∂A_1). Since this hypersurface is the level set of the function I (resp. det), its second fundamental form is the restriction of $\mathrm{D}^2I/|\mathrm{D}I|$ (resp. $\mathrm{D}^2\det/|\mathrm{D}\det|$) to the tangent space of $\Lambda_{I(A_1)}$ (resp. ∂A_1). Lemma 10.1.5 gives

$$\frac{\mathrm{D}^2I(x)}{|\mathrm{D}I(x)|}=\frac{(\alpha-1)\sum_{1\leqslant i,j\leqslant n}|x_{i,j}|^{\alpha-2}E^{i,j}\otimes E^{i,j}}{\bigl(\sum_{1\leqslant i,j\leqslant n}|x_{i,j}|^{2(\alpha-1)}\bigr)^{1/2}} = \frac{\alpha-1}{|x|_{2(\alpha-1)}^{\alpha-1}}\sum_{1\leqslant i,j\leqslant n}|x_{i,j}|^{\alpha-2}E^{i,j}\otimes E^{i,j}\,.$$

To calculate the second fundamental form of ∂A_1 at a point x in $\mathcal{D}_{A_1}$, notice that for h tangent to ∂A_1, Lemma 10.0.1 implies $\mathrm{tr}(x^{-1}h)=0$. Consequently, for h,k in $T_x\partial A_1$, Lemma 10.0.1 yields $\mathrm{D}^2\det(x)(h,k)=-\mathrm{tr}(x^{-1}hx^{-1}k)$. Then, Lemma 10.0.2 and (10.0.1) show that for $x\in\mathcal{D}_{A_1}$, the matrix $\mathrm{D}^2\det(x)/|\mathrm{D}\det(x)|$ is

$$-\sum_{1\leqslant i,j,l,m\leqslant n}(x^{-1})_{m,i}(x^{-1})_{j,l}E^{i,j}\otimes E^{l,m}/|(x^{-1})^{\mathrm{T}}|$$

and the result follows. ∎

10.1.8. LEMMA. *If $\alpha=2$ and x belongs to $\mathcal{D}_{A_1}=\mathrm{SO}(n,\mathbb{R})$, then*

$$G_{A_1}(x)=\frac{2}{\sqrt{n}}\mathrm{Id}_{\mathbb{R}^{(n-1)(n+2)/2}}\,.$$

Proof. For $\alpha=2$ the differential of I is the identity. If x is in $\mathcal{D}_{A_1}$, Lemma 10.1.2 forces $|\mathrm{D}I(x)|=|x|=\sqrt{n}$. Furthermore,

since $\mathcal{D}_{A_1}$ is the special orthogonal group, $\mathrm{Ddet}(x) = (x^{-1})^{\mathrm{T}} = x$ on $\mathcal{D}_{A_1} = \mathrm{SO}(n, \mathbb{R})$. Consequently, $|\mathrm{Ddet}(x)| = \sqrt{n}$ on $\mathcal{D}_{A_1}$.

If h, k are in $T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1}$, and x is in $\mathrm{SO}(n,\mathbb{R})$, Lemma 10.1.6 implies $(x^{-1}h)^{\mathrm{T}} = x^{-1}h$ since $x^{-1}h$ is symmetric as well as $\mathrm{tr}(x^{-1}h) = 0$. We then infer from Lemma 10.0.1 that

$$
\begin{aligned}
\mathrm{D}^2\det(x)(h,k) &= -\mathrm{tr}(x^{-1}hx^{-1}k) = -\mathrm{tr}\big((x^{-1}h)^{\mathrm{T}}x^{-1}k\big) = -\mathrm{tr}(h^{\mathrm{T}}k) \\
&= -\langle h\,, k\rangle\,.
\end{aligned}
$$

Consequently, the restriction of the bilinear form $\mathrm{D}^2\det$ to $T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1}$ is the identity. When x is in $\mathcal{D}_{A_1}$, it follows that $G_{A_1}(x)(h\,,k) = 2\langle h,k\rangle/\sqrt{n}$ on $T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1}$. Since $T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1}$ has dimension $(n-1)(n+2)/2$ thanks to Lemma 10.1.6, the result follows. ∎

In order to obtain an expression for G_{A_1} when $\alpha \neq 2$, we find an explicit orthonormal basis of $T_x\Lambda_{I(A_1)} \ominus T_x\mathcal{D}_{A_1}$ for x in $\mathcal{D}_{A_1}$. It is then possible to express the matrix G_{A_1} in this basis. The construction goes as follows.

For a real matrix $M = (M_{i,j})_{1\leqslant i,j\leqslant n}$ we denote by $M_{\bullet,j} = (M_{i,j})_{1\leqslant i\leqslant n}$ (resp. $M_{i,\bullet}$) the vector in $\mathbb{R}^n$ made of its j-th column (resp. i-th row).

Notice that any x in $\mathcal{D}_{A_1}$ is also in $\mathrm{SL}(n,\mathbb{R})$, and so is invertible. For $1 \leqslant i,j \leqslant n$, $i \neq j$, let $y_j^i \in \mathbb{R}^n$ be an orthonormal basis of $\{\,(x^{-1})_{\bullet,i}\,\}^{\perp}$, the orthogonal subspace in $\mathbb{R}^n$ of the i-th column vector of x^{-1}. Define also $e = \big((x^{-1\mathrm{T}}x^{-1})_{i,i}\big)_{1\leqslant i\leqslant n} \in \mathbb{R}^n$, the vector whose coordinates are the diagonal entries of $x^{-1\mathrm{T}}x^{-1}$. Furthermore, define y_i^i, $1 \leqslant i \leqslant n-1$ to be an orthonormal basis of $\{\,e\,\}^{\perp}$ where $\{\,e\,\}^{\perp}$ — in $\mathbb{R}^n$ — is equipped with the quadratic form $\mathrm{Proj}_{\{e\}^\perp}\mathrm{diag}\big((x^{-1\mathrm{T}}x^{-1})_{i,i}\big)_{1\leqslant i\leqslant n}\big|_{\{\,e\,\}^{\perp}}$. This quadratic form is the compression to $\{\,e\,\}^{\perp}$ of the diagonal matrix obtained by writing the components of e on its diagonal. We denote by $y_{k,j}^i$, $1 \leqslant k \leqslant n$, the components of the vector y_j^i in $\mathbb{R}^n$, $1 \leqslant i,j \leqslant n$. Finally, define

$$
\begin{aligned}
F^{i,j} &= \sum_{1\leqslant k\leqslant n} y_{k,j}^i E^{i,k}\,, \qquad i\neq j\,,\ 1\leqslant i,j\leqslant n\,, \\
F^{i,i} &= \sum_{1\leqslant l,m\leqslant n} y_{l,i}^i (x^{-1})_{m,l} E^{l,m}\,, \qquad 1\leqslant i\leqslant n-1\,.
\end{aligned}
$$

10.1.9. LEMMA. *In $\mathrm{M}(n,\mathbb{R}) \equiv \mathbb{R}^{n^2}$, the n^2-1 vectors $F^{i,j}$, for i,j in $\{\,1,2,\ldots,n\,\}$ with $(i,j)\neq(n,n)$, form an orthonormal basis of $\{\,x^{-1\mathrm{T}}\,\}^{\perp}$.*

Proof. Let us first show that all the matrices $F^{i,j}$ are orthogonal to $x^{-1\mathrm{T}}$. Indeed, if $i \neq j$,

$$\langle F^{i,j}, x^{-1\mathrm{T}} \rangle = \sum_{1 \leqslant k \leqslant n} y^i_{k,j} (x^{-1})_{k,i} = \langle y^i_j, (x^{-1})_{\bullet,i} \rangle = 0\,,$$

while for $i = j$,

$$\begin{aligned}\langle F^{i,i}, x^{-1\mathrm{T}} \rangle &= \sum_{1 \leqslant l,m \leqslant n} y^i_{l,i} (x^{-1})_{m,l} (x^{-1})_{m,l} = \sum_{1 \leqslant l \leqslant n} y^i_{l,i} (x^{-1\mathrm{T}} x^{-1})_{l,l} \\ &= \langle y^i_i \,, e \rangle = 0\,.\end{aligned}$$

To check that we have an orthonormal basis, we use the identity $\langle E^{i,j}, E^{k,l} \rangle = \delta_{(i,j),(k,l)}$. If $i \neq j$ and $p \neq q$,

$$\langle F^{i,j}, F^{p,q} \rangle = \sum_{1 \leqslant k,l \leqslant n} y^i_{k,j} y^p_{k,q} \delta_{i,p} = \langle y^i_j \,, y^p_q \rangle \delta_{i,p} = \delta_{(i,j),(p,q)}\,.$$

Next, for $i \neq j$,

$$\langle F^{i,j}, F^{p,p} \rangle = \sum_{1 \leqslant k \leqslant n} y^i_{k,j} y^p_{i,p} (x^{-1})_{k,i} = y^p_{i,p} \langle y^i_j \,, (x^{-1})_{\bullet,i} \rangle = 0\,.$$

Finally, for $i, j = 1, 2, \ldots, n-1$,

$$\begin{aligned}\langle F^{i,i}, F^{j,j} \rangle &= \sum_{1 \leqslant l,m \leqslant n} y^i_{l,i} y^j_{l,j} (x^{-1})_{m,l} (x^{-1})_{m,l} \\ &= \sum_{1 \leqslant l \leqslant n} y^i_{l,i} y^j_{l,j} (x^{-1\mathrm{T}} x^{-1})_{l,l} \\ &= \delta_{i,j}\end{aligned}$$

by our choice of y^i_i. ∎

Combining Lemmas 10.1.7 and 10.1.9, we can calculate the $(n^2 - 1) \times (n^2 - 1)$-matrix $\big(\langle G_{A_1}(x) F^{i,j}, F^{k,l} \rangle\big)$ where (i, j), (k, l) belong to $\{\,1, \ldots, n\,\} \times \{\,1, \ldots, n\,\} \setminus \{\,(n, n)\,\}$. This amounts to writing the matrix $G_{A_1}(x)$ in the orthonormal basis $F^{i,j}$, $1 \leqslant i, j \leqslant n$, $(i, j) \neq (n, n)$. The explicit calculation is rather long, and unfortunately does not seem to simplify much. But the work done so far is all that we need to implement the approximation numerically.

To conclude the proof of Theorem 10.1.1, it remains to check the assumptions of Theorem 7.1.

Assumptions (7.1) and (7.2) hold since we assume $\alpha > 1$. Assumption (7.3) is trivial since we assume $\alpha \geqslant 2$.

Assumption (7.4) is guaranteed by Conjecture 10.1.3 when $\alpha \neq 2$, while it is trivial for $\alpha = 2$.

To check assumption (7.5), use Remark 7.3 and the calculation of Ddet and DI made in this section. Indeed, Lemma 10.1.1 yields

$$\langle \mathrm{D}I(x), \mathrm{Ddet}(x)\rangle = \frac{1}{\lambda}|x|_{2(\alpha-1)}^{2(\alpha-1)} > 0$$

for any x in $\mathcal{D}_{A_1}$. This concludes the proof of Theorem 10.1.4. ■

When $\alpha = 2$, we infer the following corollary.

10.1.10. COROLLARY. *Let $(X_{i,j})_{1\leqslant i,j\leqslant n}$ be an $n\times n$ random matrix, with independent and identically coefficients all having a standard normal distribution. The distribution of $t^{-1/n}X$ given* $\mathrm{det}X \geqslant t$ *converges weakly* to the uniform distribution over* $\mathrm{SO}(n,\mathbb{R})$.

Proof. It is now a straightforward application of Theorem 7.5. ■

10.2. Random determinants, heavy tails.

We now consider the problem of approximating the tail probability of the determinant of a random matrix, assuming that its coefficients are independent and all have a Student like cumulative distribution function S_α. For this problem, the general framework proposed so far can be used. The argument is very much like that used to prove Theorem 8.3.1. For a change, we will give a probabilistic proof, which is actually inspired by Theorem 5.1, showing another sort of use of that theorem. This proof will be far less conceptual, and will give no insights.

The result is as follows.

10.2.1. THEOREM. *Let $X = (X_{i,j})_{1\leqslant i,j\leqslant n}$ be a matrix with independent and identically distributed coefficients, all having a Student like distribution with parameter α. Then,*

$$P\{\,\mathrm{det}X \geqslant t\,\} \sim \frac{n(2K_{s,\alpha}\alpha^{(\alpha+1)/2})^n}{2\alpha}\frac{(\log t)^{n-1}}{t^\alpha} \qquad as\ t\to\infty\,.$$

Proof. We will see why Theorem 5.1 suggests that, as t tends to infinity,

$$\begin{aligned} P\{\,\det X \geqslant t\,\} &= P\Big\{ \sum_{\sigma\in\mathfrak{S}_n} \epsilon(\sigma) \prod_{1\leqslant i\leqslant n} X_{i,\sigma(i)} \geqslant t \Big\} \\ &\sim \sum_{\sigma\in\mathfrak{S}_n} P\Big\{ \epsilon(\sigma) \prod_{1\leqslant i\leqslant n} X_{i,\sigma(i)} \geqslant t \Big\} \\ &= n!P\Big\{ \prod_{1\leqslant i\leqslant n} X_{1,i} \geqslant t \Big\}\,. \end{aligned}$$

Admitting this relation, our first lemma gives the key estimate.

10.2.2. LEMMA. *Let $X_1,\ldots,X_n$ be n independent random variables with Student-like distribution S_α. The product $X_1\ldots X_n$ has upper tail*

$$P\{\,X_1\ldots X_n \geqslant t\,\} \sim \frac{(2K_{s,\alpha}\alpha^{\frac{\alpha+1}{2}})^n}{2\alpha(n-1)!}\frac{(\log t)^{n-1}}{t^\alpha} \qquad \textit{as } t\to\infty\,.$$

Its lower tail is equivalent to its upper tail.

Proof. We proceed by induction. For $n=1$, the result is plain from the definition of Student-like distributions. Call F_n the cumulative distribution function of the product $X_1\ldots X_n$, and c_n the constant

$$c_n = \frac{(2K_{s,\alpha}\alpha^{(\alpha+1)/2})^n}{2\alpha(n-1)!}\,.$$

Assume that F_{n-1} has the form given in the statement of the lemma. Then

$$\begin{aligned} 1-F_n(t) = P\{\,X_1\ldots X_{n-1} \geqslant t/X_n\,;\,X_n>0\,\} \\ + P\{\,X_1\ldots X_{n-1}\leqslant t/n\,;\,X_n<0\,\}\,. \end{aligned}$$

Let us evaluate the first probability in the sum. The second one is either evaluated in the same way, or is obtained from the first one by changing X_n into $-X_n$. This first probability can be rewritten as $\int_0^\infty 1-F_{n-1}(t/x)\,\mathrm{d}S_\alpha(x)$.

Let δ be a positive number. Using the induction hypothesis, there exists a positive M such that for any y larger than M,

$$(1-\delta)c_{n-1}\frac{(\log y)^{n-2}}{y^\alpha} \leqslant 1-F_{n-1}(y) \leqslant (1+\delta)c_{n-1}\frac{(\log y)^{n-2}}{y^\alpha}\,.$$

Moreover, taking M large enough, we also have

$$(1-\delta)\frac{c_1}{y^\alpha} \leqslant 1 - S_\alpha(y) \leqslant (1+\delta)\frac{c_1}{y^\alpha}\,.$$

Consequently,

$$\int_M^{t/M} 1 - F_{n-1}(t/x)\,\mathrm{d}S_\alpha(x) \leqslant (1+\delta)c_{n-1}\int_M^{t/M} \frac{(\log t/x)^{n-2}}{(t/x)^\alpha}\,\mathrm{d}S_\alpha(x)\,. \tag{10.2.1}$$

We integrate by parts, writing

$$\int_M^{t/M} x^\alpha(\log t/x)^{n-2}\,\mathrm{d}S_\alpha(x) = \Bigl[x^\alpha(\log t/x)^{n-2}\bigl(S_\alpha(x)-1\bigr)\Bigr]_M^{t/M}$$
$$-\int_M^{t/M} \bigl(\alpha x^{\alpha-1}(\log t/x)^{n-2} - (n-2)x^{\alpha-1}(\log t/x)^{n-3}\bigr)\bigl(S_\alpha(x)-1\bigr)\,\mathrm{d}x\,.$$

The number M can be taken large enough so that $1/\log z \leqslant \delta$ for any z greater than M. Then,

$$\begin{aligned}
&\int_M^{t/M} x^\alpha(\log t/x)^{n-2}\mathrm{d}S_\alpha(x)\\
&\qquad\leqslant O(\log t)^{n-2} + \bigl(\alpha(1+\delta)c_1 + \delta(n-2)\bigr)\int_M^{t/M}\frac{1}{x}\Bigl(\log\frac{t}{x}\Bigr)^{n-2}\mathrm{d}x\\
&\qquad= O(\log t)^{n-2} + \bigl(\alpha c_1(1+\delta) + \delta(n-2)\bigr)\int_M^{t/M}\frac{1}{y}(\log y)^{n-2}\mathrm{d}y\\
&\qquad= O(\log t)^{n-2} + \bigl(\alpha c_1(1+\delta) + \delta(n-2)\bigr)\frac{(\log t)^{n-1}}{n-1}\bigl(1+o(1)\bigr)
\end{aligned}$$

as t tends to infinity. Therefore, (10.3.1) yields

$$\begin{aligned}
&\int_M^{t/M} 1 - F_{n-1}(t/x)\,\mathrm{d}S_\alpha(x)\\
&\qquad\leqslant \frac{(1+\delta)}{t^\alpha}c_{n-1}\bigl(\alpha(1+\delta)c_1 + \delta(n-2)\bigr)\frac{(\log t)^{n-1}}{n-1}\bigl(1+o(1)\bigr)\,.
\end{aligned}$$

We obtain a similar lower bound, replacing δ by $-\delta$.

In the range of integration $x > t/M$, we have

$$\int_{t/M}^\infty 1 - F_{n-1}(t/x)\,\mathrm{d}S_\alpha(x) \leqslant 1 - S_\alpha(t/M) = O(t^{-\alpha}) \qquad \text{as } t\to\infty\,.$$

On the other hand, when $x < M$ we have

$$\int_0^M \big(1 - F_{n-1}(t/x)\big)\,\mathrm{d}S_\alpha(x) \leqslant 1 - F_{n-1}(t/M) = O\Big(\frac{(\log t)^{n-2}}{t^\alpha}\Big)$$

as t tends to infinity. Since δ is arbitrary, we proved that

$$P\{\, X_1 \dots X_{n-1} \geqslant t/X_n\,;\, X_n \geqslant 0 \,\} \sim \alpha \frac{c_1 c_{n-1}}{n-1} \frac{(\log t)^{n-1}}{t^\alpha}\,,$$

as t tends to infinity. Therefore,

$$1 - F_n(t) \sim 2\alpha \frac{c_1 c_{n-1}}{n-1} \frac{(\log t)^{n-1}}{t^\alpha}\,.$$

Since $2\alpha c_1 c_{n-1} = c_n$, the result on the upper tail follows. The lower tail $F(t)$ as t tends to infinity is handled in the same way. ∎

The next lemma will allow us to prove that if $\det X \geqslant t$ and t is large, it is very unlikely that two different products $\prod_{1\leqslant i\leqslant n} X_{i,\sigma(i)}$, $\sigma \in \mathfrak{S}_n$, are both of order t.

10.2.3. LEMMA. *Let $X_1, \dots, X_{n+k}$ be $n+k$ independent random variables with Student-like distribution function S_α. For $\alpha > 1$, and k positive,*

$$P\{\, X_1 \dots X_n \geqslant t\,;\, X_{k+1} \dots X_{k+n} \geqslant t \,\} = o\Big(\frac{(\log t)^{n-1}}{t^\alpha}\Big)$$

as t tends to infinity.

Proof. If $k \geqslant n$, the result follows from independence and Lemma 10.2.2. Thus, from now on, assume that $k < n$. Set $Y = X_1 \dots X_k$, $Z = X_{k+1} \dots X_n$ and $U = X_{n+1} \dots X_{n+k}$. Since these random variables are independent, we have

$$\begin{aligned} P\{\, YZ \geqslant t; ZU \geqslant t \,\} &= \int_{y,u\geqslant 0} P\Big\{\, Z \geqslant t\Big(\frac{1}{y} \vee \frac{1}{u}\Big) \,\Big\}\,\mathrm{d}F_k(y)\mathrm{d}F_k(u) \\ &\quad + \int_{y,u\leqslant 0} P\Big\{\, Z \leqslant t\Big(\frac{1}{y} \wedge \frac{1}{u}\Big) \,\Big\}\,\mathrm{d}F_k(y)\mathrm{d}F_k(u)\,. \end{aligned}$$

Let us evaluate the first integral, the second one being similar. Using the symmetry in u and y, it suffices to prove that

$$\int_{0\leqslant y\leqslant u} P\{\, X \geqslant t/y \,\}\,\mathrm{d}F_k(y)\mathrm{d}F_k(u) = o\big(t^{-\alpha}(\log t)^{n-1}\big)$$

as t tends to infinity.

Let us use the notation c_k as in the proof of Lemma 10.2.2. Let δ be an arbitrary positive number. Then, there exists a positive M such that for any $u > M$

$$(1-\delta)c_{n-k}\frac{(\log u)^{n-k-1}}{u^\alpha} \leqslant P\{\, Z \geqslant u \,\} \leqslant (1+\delta)c_{n-k}\frac{(\log u)^{n-k-1}}{u^\alpha}$$

$$(1-\delta)c_k\frac{(\log u)^{k-1}}{u^\alpha} \leqslant 1-F_k(u) \leqslant (1+\delta)c_k\frac{(\log u)^{k-1}}{u^\alpha}\,.$$

We then have, using Lemma 10.2.2 and the fact that k is strictly less than n

$$\begin{aligned}\int_{\substack{0<y<u\\ t/M<u}} & P\{\, Z \geqslant t/y \,\}\,\mathrm{d}F_k(y)\,\mathrm{d}F_k(u)\\ &\leqslant \int_{t<M<u} \mathrm{d}F_k(u) = O\big(t^{-\alpha}(\log t)^{k+1}\big)\\ &= o\big(t^{-\alpha}(\log t)^{n-1)}\big)\,.\end{aligned}$$

Thus, we need to prove that

$$\int_{M<y<u<t/M} P\{\, Z \geqslant t/y \,\}\,\mathrm{d}F_k(y)\mathrm{d}F_k(u) = o\big(t^{-\alpha}(\log t)^{n-1}\big) \tag{10.2.2}$$

as t tends to infinity. We first perform the integration in u, obtaining

$$\begin{aligned}&\int_{M<y<u<t/M} P\{\, Z \geqslant t/y \,\}\,\mathrm{d}F_k(u)\,\mathrm{d}F_k(y)\\ &= \int_{M<y<t/M} P\{\, Z \geqslant t/y \,\}\big(1-F_k(y)-1+F_k(t/M)\big)\,\mathrm{d}F_k(y)\\ &\leqslant \big(1-F_k(M)\big)\int_{M<y<t/M} P\{\, Z \geqslant t/y \,\}\,d\,F_k(y)\,.\end{aligned}$$

We then use the bound on the tail of Z and integrate by parts,

$$\begin{aligned}\int_{M<y<t/M} & P\{\, Z \geqslant t/y \,\}\,\mathrm{d}F_k(y)\\ &\leqslant (1+\delta)c_{n-k}\int_{M<y<t/M}\Big(\log\frac{t}{y}\Big)^{n-k-1}\Big(\frac{y^\alpha}{t^\alpha}\,\mathrm{d}F_k(y)\\ &\leqslant (1+\delta)c_{n-k}\Big[\Big(\log\frac{t}{y}\Big)^{n-k-1}\Big(\frac{y^\alpha}{t^\alpha}\big(F_k(y)-1\big)\Big]_M^{t/M}\\ &\qquad + (1+\delta)c_{n-k}\int_{M<y<t/M}\frac{y^{\alpha-1}}{t^\alpha}\Big(\log\frac{t}{y}\Big)^{n-k-2}\times\\ &\qquad\quad \Big(\alpha\log\frac{t}{y}+(n-k-1)\Big)\big(F_k(y)-1\big)\,\mathrm{d}y\end{aligned}$$

Using the bound on $1-F_k$, we obtain

$$\begin{aligned}\int_{M<y<t/M} & P\{\, Z \geqslant t/y \,\} \,\mathrm{d}F_k(y) \\ & \leqslant O\big((\log t)^{n-k-1}/t^\alpha\big) \\ & \quad + O(1)\int_{M<y<t/M} \frac{(\log t/y)^{n-k-1}}{t^\alpha}\frac{(\log y)^{k-1}}{y}\,\mathrm{d}y\end{aligned}$$

The change of variable $v=(\log y)/\log t$ shows that

$$\begin{aligned}\int_{M<y<t/M} & \frac{(\log t-\log y)^{n-k-1}}{t^\alpha}\frac{(\log y)^{k-1}}{y}\,\mathrm{d}y \\ & \sim \frac{(\log t)^{n-2}}{t^\alpha}\int_0^1 (1-v)^{n-k-1}v^{k-1}\,\mathrm{d}v \\ & = o\Big(\frac{(\log t)^{n-1}}{t^\alpha}\Big)\,.\end{aligned}$$

Consequently, (10.2.2) holds as well as Lemma 10.2.3. ∎

We can now prove Theorem 10.2.1. For a permutation σ in $\mathfrak{S}_n$, define

$$Y_\sigma = \epsilon(\sigma)\prod_{1\leqslant i\leqslant n} X_{i,\sigma(i)}\,.$$

For any fixed positive δ,

$$\begin{aligned}P\{\, \mathrm{det}X \geqslant t \,\} &= P\Big\{\, \sum_{\sigma\in\mathfrak{S}_n} Y_\sigma \geqslant t \,\Big\} \\ &\geqslant P\Big\{ \bigcup_{\sigma\in\mathfrak{S}_n} \Big(\{\, Y_\sigma \geqslant t(1+\delta) \,\} \cap \bigcap_{\tau\in\mathfrak{S}_n\setminus\{\sigma\}} \{\, |Y_\tau| \leqslant t\delta/n! \,\} \Big) \Big\} \\ &\geqslant \sum_{\sigma\in\mathfrak{S}_n} P\Big\{ Y_\sigma \geqslant t(1+\delta) \,\} \cap \bigcap_{\tau\in\mathfrak{S}_n\setminus\{\sigma\}} \{\, |Y_\tau| \leqslant t\delta/n! \,\Big\} \\ &\quad - \sum_{\substack{\sigma_1,\sigma_2\in\mathfrak{S}_n \\ \sigma_1\neq\sigma_2}} P\{\, Y_{\sigma_1} \geqslant t(1+\delta)\,;\, Y_{\sigma_2} \geqslant t(1+\delta) \,\}\end{aligned}$$

From Lemma 10.2.3, we infer that

$$\sum_{\substack{\sigma_1,\sigma_2\in\mathfrak{S}_n \\ \sigma_1\neq\sigma_2}} P\{\, Y_{\sigma_1} \geqslant t(1+\delta)\,;\, Y_{\sigma_2} \geqslant t(1+\delta) \,\} = o\Big(\frac{(\log t)^{n-1}}{t^\alpha}\Big)\,.$$

Moreover, if τ and σ are distinct, Lemma 10.2.3 implies

$$P\{\, Y_\sigma \geqslant t(1+\delta) \text{ and } |Y_\tau| \geqslant t/\delta n! \,\} = o\Big(\frac{(\log t)^{n-1}}{t^\alpha}\Big)\,.$$

Consequently, using Lemma 10.2.2,

$$\begin{aligned}\sum_{\sigma\in\mathfrak{S}_n} P\Big\{\, Y_\sigma \geqslant t(1+\delta)\,\} \cap \bigcap_{\tau\in\mathfrak{S}_n\setminus\{\sigma\}} \{\, |Y_\tau| \leqslant t\delta/n!\,\}\Big\}&\\ \sim n!P\{\, Y_{\rm Id} \geqslant t(1+\delta)\,\}&\\ \sim \frac{n}{2\alpha}(2K_{s,\alpha}\alpha^{\frac{\alpha+1}{2}})^n \frac{(\log t)^{n-1}}{\big((1+\delta)t\big)^{\alpha+1}} \qquad &\text{as } t\to\infty\,.\end{aligned}$$

This proves the lower bound

$$P\{\,\det X \geqslant t\,\} \geqslant \frac{n}{2\alpha}(2K_{s,\alpha}\alpha^{\frac{\alpha+1}{2}})^n \frac{(\log t)^{n-1}}{t^{\alpha+1}}\frac{1+o(1)}{(1+\delta)^{\alpha+1}}$$

as t tends to infinity.

To obtain a matching upper bound, notice that

$$\begin{aligned}P\{\,\det X \geqslant t\,\} \leqslant P\{\, &\exists\,\sigma\in\mathfrak{S}_n\,,\, Y_\sigma \geqslant t(1-\delta)\\ &\text{and } \forall\tau\in\mathfrak{S}_n\setminus\{\sigma\}\,,\, |Y_\sigma| \leqslant t\delta/n!\,\}\\ +P\{\, &\exists\,\tau_1,\tau_2\in\mathfrak{S}_n\,,\, \tau_1\neq\tau_2\,,\, Y_{\tau_1}\geqslant \delta t/n!\,;\, Y_{\tau_2}\geqslant \delta t/n!\,\}\end{aligned}$$

Applying Lemma 10.2.3 and 10.2.2, we obtain

$$\begin{aligned}P\{\,\det X \geqslant t\,\} &\leqslant \sum_{\sigma\in\mathfrak{S}_n} P\{\, Y_\sigma \geqslant t(1-\delta)\,\} + o\Big(\frac{(\log t)^{n-1}}{t^\alpha}\Big)\\ &\sim n\frac{(2K_{s,\alpha}\alpha^{\frac{\alpha+1}{2}})^n}{2\alpha}\frac{(\log t)^{n-1}}{t^\alpha}\frac{1+o(1)}{(1-\delta)^\alpha} \qquad \text{as } t\to\infty\,.\end{aligned}$$

Since δ is arbitrarily small, we proved Theorem 10.2.1. ∎

Let us now show why Theorem 5.1 suggested the proof of Theorem 10.2.1. Define the set

$$A_t = \{\, x\in \mathrm{M}(n,\mathbb{R}) \,:\, \det x \geqslant t\,\} = t^{1/n}A_1\,.$$

Theorem 10.2.1 provides an estimate for the integral

$$\int_{A_t} \prod_{1\leqslant i,j\leqslant n} s_\alpha(x_{i,j})\,\mathrm{d}x_{i,j}\,.$$

The change of variable $Y = \Phi^{\leftarrow} \circ S_\alpha(X)$ leads us to introduce

$$B_t = \Phi^{\leftarrow} \circ S_\alpha(A_t)\,.$$

It allows us to rewrite the integral under consideration as

$$\int_{B_t} e^{-I(y)}\,\mathrm{d}y\,,$$

where

$$I(y) = \frac{|y|^2}{2} + \log(2\pi)^{n^2}$$

is convex. To minimize I over B_t, take a matrix x in ∂A_1 that is in $\mathrm{SL}(n,\mathbb{R})$. Then $y = \Phi^{\leftarrow} \circ S_\alpha(t^{1/n}x)$ is on the boundary of B_t. Furthermore,

$$I\big(\Phi^{\leftarrow} \circ S_\alpha(t^{1/n}x)\big) \sim (\log t^{1/n})\sharp\{\,(i,j) \,:\, x_{i,j} \neq 0\,\}\,.$$

Thus, for $I\big(\Phi^{\leftarrow} \circ S_\alpha(t^{1/n}x)\big)$ to be minimum asymptotically, x should have as many zero components as possible, namely n. The matrices of $\mathrm{SL}(n,\mathbb{R})$ with n nonvanishing entries form a subgroup which can be described as follows. Define the matrix

$$I_{1,n-1} = \begin{pmatrix} -1 & 0 \\ 0 & \mathrm{Id}_{n-1} \end{pmatrix}\,.$$

Let $\mathrm{DSL}(n,\mathbb{R})$ be the subgroup of all diagonal matrices in $\mathrm{SL}(n,\mathbb{R})$. To a permutation σ in $\mathfrak{S}_n$ we associate the matrix of its permutation representation, conveniently denoted σ as well. Thus, $\sigma e_i = e_{\sigma(i)}$. Denote by $\mathfrak{S}_{n,+}$ the subgroup of all even permutation of n elements. Equivalently, $\mathfrak{S}_{n,+}$ is $\mathfrak{S}_n \bigcap \mathrm{SL}(n,\mathbb{R})$. Denote $\mathfrak{S}_{n,-}$ the subset of $\mathfrak{S}_n$ of all odd permutation matrices. Let $\langle I_{1,n-1}\mathfrak{S}_{n,-}\rangle$ be the subgroup of $\mathrm{SL}(n,\mathbb{R})$ generated by the matrices $I_{1,n-1}\sigma$, with $\sigma \in \mathfrak{S}_{n,-}$. Then $\mathfrak{S}_{n,+} \cup \langle I_{1,n-1}\mathfrak{S}_{n,-}\rangle$ is a group made of matrices which are, up to the sign of their entries, permutation matrices, and are of determinant equal to 1. This group acts on $\mathrm{DSL}(n,\mathbb{R})$ by

$$(\sigma, m) \in \big(\mathfrak{S}_{n,+} \cup \langle I_{1,n-1}\mathfrak{S}_{n,-}\rangle\big) \times \mathrm{DSL}(n,\mathbb{R}) \mapsto \sigma m \in \mathrm{SL}(n,\mathbb{R})$$

Denote by $\big(\mathfrak{S}_{n,+} \cup \langle I_{1,n-1}\mathfrak{S}_{n,-}\rangle\big)\mathrm{DSL}(n,\mathbb{R})$ the image of this action. One easily sees that it is a subgroup of $\mathrm{SL}(n,\mathbb{R})$, made of all the matrices with exactly n nonvanishing entries. Let $x = \sigma m$ be

in this subgroup. Using Lemma A.1.5 and the fact that $\det m = \prod_{1\leqslant i\leqslant d} m_{i,i} = 1$,

$$\begin{aligned}
I\big(\Phi^{\leftarrow}\circ S_\alpha(t^{1/\alpha}x)\big) &= I\big(\Phi^{\leftarrow}\circ S_\alpha(t^{1/d}m)\big) \\
&= \sum_{1\leqslant i\leqslant d}\Big(\alpha\log(t^{1/n}|m_{i,i}|\big) - \frac{1}{2}\log\log(t^{1/n}|m_{i,i}|) \\
&\qquad - 2\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})\Big) + \log(2\pi)^{n^2/2} + o(1) \\
&= \alpha\log t - \frac{d}{2}\log\log t^{1/d} - 2d\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) \\
&\qquad + \log(2\pi)^{n^2/2} + o(1)\,. \qquad (10.2.3)
\end{aligned}$$

This expression does not depends on m. It suggests that the dominating manifold in our problem should be

$$\Phi^{\leftarrow}\circ S_\alpha\Big(t^{1/n}\big(\mathfrak{S}_{n,+}\cup\langle I_{1,n-1}\mathfrak{S}_{n,-}\rangle\big)\mathrm{DSL}(n,\mathbb{R})\Big)\,.$$

This set is made up of $n!$ connected components, each component being $\mathrm{DSL}(n,\mathbb{R})$ composed on the left either by an even permutation, or by $I_{1,n-1}$ and an odd permutation. As I is invariant under permutations and composition by $I_{1,n-1}$, all these components should be equally likely. Since the distribution of the X_i's is asymptotically symmetric, this suggests the approximation

$$P\Big\{\sum_{\sigma\in\mathfrak{S}_n}\epsilon(\sigma)\prod_{1\leqslant i\leqslant n}X_{i,\sigma(i)}\geqslant t\Big\}\sim\sum_{\sigma\in\mathfrak{S}_n}P\Big\{\prod_{1\leqslant i\leqslant n}X_{i,\sigma\circ\tau(i)}\geqslant t\Big\}\,,$$

where τ is a transposition, depending on σ, such that $\mathfrak{S}_n = \{\,\sigma,\sigma\circ\tau : \sigma\in\mathfrak{S}_{n,+}\,\}$. The main reason the proof is complicated using this method is that (10.2.3) is not uniform in m. It is uniform in the range $t^{1/n}m\to\infty$ and $\log|m_{i,i}|/\log t\to 0$. This is exactly the same problem as the one we faced in section 8.3, and a similar parameterization can be used.

10.3. Geometry of the unit ball of $\mathrm{M}(n,\mathbb{R})$.

The purpose of this section is to study some elementary differential geometric properties of the set $\mathcal{S}$ of all real matrices of norm 1. This will be instrumental in the next section to obtain results on norm of random matrices. We will prove — Propositions 10.3.1 and 10.3.2 — that this set is a fiber bundle over a Klein bottle of dimension $2(n-1)$,

whose fibers are isomorphic to the unit ball of $(n-1)\times(n-1)$ real matrices. We will explicitly calculate various curvatures of this set.

Recall that the set $\mathrm{M}(n,\mathbb{R})=\mathbb{R}^{n^2}$ of all $n\times n$ matrices with real entries is equipped with the inner product

$$\langle M,N\rangle=\sum_{1\leqslant i,j\leqslant n}M_{i,j}N_{i,j}=\mathrm{tr}(MN^{\mathrm{T}})\,.$$

With this inner product, $\mathrm{M}(n,\mathbb{R})$ is $\mathbb{R}^{n^2}$ equipped with its standard inner product. On $\mathbb{R}^{n^2}$, the Euclidean unit sphere is a submanifold of dimension n^2-1 with constant curvature, whose geometry is very well understood. However, for algebraic reasons, it is often more convenient to equip $\mathrm{M}(n,\mathbb{R})$ with the operator norm

$$\|M\|=\sup\{\,|Mu|\,:\,u\in S_{n-1}\,\}\,,$$

where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^n$ and

$$S_{n-1}=\{\,x\in\mathbb{R}^n\,:\,|x|=1\,\}$$

is the unit sphere centered at the origin. The unit sphere centered in $(\mathrm{M}(n,\mathbb{R}),\|\cdot\|)$, namely

$$\mathcal{S}=\{\,M\in\mathrm{M}(n,\mathbb{R})\,:\,\|M\|=1\,\}\,.$$

is not as familiar as S_{n^2-1} as far as its geometry is concerned. We need to understand what $\mathcal{S}$ looks like in $\mathrm{M}(n,\mathbb{R})$ identified with $\mathbb{R}^{n^2}$.

For this purpose, for any $u,v\in\mathbb{R}^n$, define the subspace of matrices

$$H_{u,v}=\{\,h\in\mathrm{M}(n,\mathbb{R})\,:\,hu=h^{\mathrm{T}}v=0\,\}\,.$$

In what follows, vectors in $\mathbb{R}^n$ are considered as row vectors, and so if u belongs to $\mathbb{R}^n$, then u^{T} is a $1\times n$ matrix. We also use systematically the tensor product notation; if u, v are two vectors in $\mathbb{R}^n$, their tensor product is the matrix $u\otimes v=vu^{\mathrm{T}}$. This notation agrees with that used in section 10.1 when we dealt with vectors in $\mathbb{R}^{n^2}$.

To understand the geometry of $\mathcal{S}$, it is convenient to remove some singular points and define

$$\mathcal{S}^0=\{\,M\in\mathcal{S}\,:\,1\text{ is a simple eigenvalue of }M^{\mathrm{T}}M\,\}\,.$$

In $\mathrm{M}(n,\mathbb{R})$, the closure of $\mathcal{S}^o$ is $\mathcal{S}$. Proposition 10.3.1 bellow asserts that $\mathcal{S}^o$ is a smooth submanifold of $\mathbb{R}^{n^2}$. Moreover, $\mathcal{S}^0$ is a fiber bundle

over a Klein bottle $S_{n-1} \otimes S_{n-1} \equiv S_{n-1} \times S_{n-1}/\{\,\mathrm{Id}, -\mathrm{Id}\,\}$, whose fibers are isomorphic to the unit ball of H_{e_1,e_1} for the operator norm. So the dimension of the fibers is $(n-1)^2$. We will show that there are no higher dimensional convex subsets in $\mathcal{S}^o$ — this follows from the form of the curvature tensor of $\mathcal{S}$ given in Theorem 10.3.3. Each fiber is also is orthogonal to its base point in $S_{n-1} \times S_{n-1}/\{\,\mathrm{Id}, -\mathrm{Id}\,\}$.

10.3.1. PROPOSITION *Every matrix M in $\mathcal{S}$ can be written as $M = u \otimes v + h$ for some u, v in S_{n-1}, and $h \in H_{u,v}$ with $\|h\| \leqslant 1$. This decomposition satisfies the following properties:*
(i) up to the transformation $(u, v) \mapsto (-u, -v)$, it is unique if and only if 1 is a simple eigenvalue of $M^{\mathrm{T}}M$.
(ii) $H_{u,v}$ is orthogonal to $u \otimes v$ and $\dim H_{u,v} = (n-1)^2$ for all u, v in S_{n-1}.

Proof. To check that matrices of the form $M = u \otimes v + h$, with u, v in S_{n-1} and $\|h\| \leqslant 1$ are of unit norm, notice that the operator norm of such matrix is at least 1, since $Mu = v$. On the other hand, write any vector x of $\mathbb{R}^n$ as $u\langle x, u\rangle + \mathrm{Proj}_{u^\perp} x$ where $\mathrm{Proj}_{u^\perp}$ is the projection onto $\{\,u\,\}^\perp$. Then, apply M to x, use that h is a contraction and belongs to $H_{u,v}$ to obtain $|Mx|^2 \leqslant |x|^2$, and so $\|M\| \leqslant 1$.

To prove that all matrices of norm 1 are of this form, take u to be a unit eigenvector of $M^{\mathrm{T}}M$ with eigenvalue 1. This vector u is unique up to its sign if and only if 1 is a simple eigenvalue. Define $v = Mu$ and $h = M - u \otimes v$. Since

$$1 = |u| = |M^{\mathrm{T}}Mu| \leqslant |Mu| = |v| \leqslant |u| = 1\,,$$

the vector v also belong to S_{n-1}. One easily checks that h belongs to $H_{u,v}$. To see why h is a contraction, notice first that $hu = 0$. Moreover, if w is orthogonal to u, then $|hw| = |Mw| \leqslant |w|$. The uniqueness statement is then clear.

The orthogonality relation (ii) follows from $\langle u \otimes v, h\rangle = \mathrm{tr}(vu^{\mathrm{T}}h^{\mathrm{T}}) = 0$, for h belongs to $H_{u,v}$.

To obtain the dimension of $H_{u,v}$, write R_u as an orthogonal matrix mapping the first vector of the canonical basis of $\mathbb{R}^n$, say e_1, to u. Then $H_{u,v} = R_v^{\mathrm{T}} H_{e_1,e_1} R_u$. Hence, $\dim H_{u,v} = \dim H_{e_1,e_1}$. Since the equations determining H_{e_1,e_1} are

$$h_{1,1} = h_{1,2} = \ldots = h_{1,n} = 0 \text{ and } h_{1,1} = h_{2,1} = \ldots = h_{n,1} = 0\,,$$

we have $\dim H_{u,v} = (n-1)^2$ as claimed. ∎

For the unit sphere S_{n-1} in $\mathbb{R}^n$, it is an obvious fact that the tangent space at any point u is just the subspace orthogonal to u in $\mathbb{R}^n$. So one may wonder if this property has an analogue for the unit ball $\mathcal{S}$. Our next proposition shows that this is somewhat the case and gives an explicit description of the tangent spaces. This will be useful in calculating the curvature tensor of $\mathcal{S}^o$. It also proves that the fibers $H_{u,v}$ are not only orthogonal to $u \otimes v$ but also to the tangent space $T_{u\otimes v}(S_{n-1} \otimes S_{n-1})$. Hence, they point orthogonally to the base.

Notice that if h belongs to $H_{u,v}$, the image $\mathrm{Im}h = h\mathbb{R}^n$ is included in $\{ v \}^\perp = T_v S_{n-1}$, while $\mathrm{Im}h^{\mathrm{T}} \subset \{ u \}^\perp = T_u S_{n-1}$. For u, v in S_{n-1} and h in $H_{u,v}$, consider the following subspaces of $\mathrm{M}(n, \mathbb{R})$,

$$
\begin{aligned}
H^1_{u,v,h} &= \{ a \otimes v - u \otimes (ha) \, : \, a \in T_u S_{n-1} \} \,, \\
H^2_{u,v,h} &= \{ u \otimes b - (h^{\mathrm{T}} b) \otimes v \, : \, b \in T_v S_{n-1} \} \,.
\end{aligned}
$$

10.3.2. PROPOSITION. *Let u, v be in S_{n-1} and h be in $H_{u,v}$ with $\|h\| < 1$. Then*
(i) $H^1_{u,v,h} \cap H^2_{u,v,h} = \{ 0 \}$;
(ii) $T_{u\otimes v + h}\mathcal{S}^0 = H_{u,v} \oplus (H^1_{u,v,h} + H^2_{u,v,h})$.
The vector $u \otimes v$ is an outward unit normal to $\mathcal{S}^0$ at all points of the form $u \otimes v + h$, with $u, v \in S_{n-1}$ and h a contraction belonging to $H_{u,v}$.

Proof. It is convenient to notice the following trivial identity which will be used repeatedly: for any a, b, x, y in $\mathbb{R}^n$,

$$
\langle a \otimes b, x \otimes y \rangle = \mathrm{tr}(b a^{\mathrm{T}} x y^{\mathrm{T}}) = \langle a, x \rangle \langle b, y \rangle \,.
$$

(i) Let a be in $T_u S_{n-1}$, and b be in $T_v S_{n-1}$. Define

$$
\begin{aligned}
x &= a \otimes v - u \otimes (ha) \in H^1_{u,v,h} \,, \\
y &= u \otimes b - (h^{\mathrm{T}} b) \otimes v \in H^2_{u,v,h} \,.
\end{aligned}
$$

Since u is orthogonal to a and v to b, and h is in $H_{u,v}$,

$$
|\langle x, y \rangle| = |\langle ha, b \rangle + \langle a, h^{\mathrm{T}} b \rangle| \leqslant |ha||b| + |a||h^{\mathrm{T}} b| \,.
$$

Moreover, for the same reasons,

$$
|x|^2 = |a|^2 + |ha|^2 \,, \qquad \text{and} \qquad |y|^2 = |b|^2 + |h^{\mathrm{T}} b|^2 \,.
$$

Therefore $|\langle x, y\rangle| < |x||y|$. Thus, x and y cannot be collinear, and $H^1_{u,v,h} \cap H^2_{u,v,h} = \{\, 0 \,\}$.
(ii) The inclusion of $H_{u,v}$ into $T_{u\otimes v+h}\mathcal{S}^0$ is clear: consider the tangent vector at 0 of the curve $s \mapsto u \otimes v + (1+s)h \in \mathcal{S}^0$.

Next, consider two curves $u(s)$, $v(s)$ in S_{n-1}, with $u(0) = u$, $v(0) = v$, $u'(0) = a$, $v'(0) = b$. Let $h(s)$ be a curve in $\mathrm{M}(n, \mathbb{R})$ such that $h(s)$ is in $H_{u(s),v(s)}$, $h(0) = h$, and $h'(0) = k$. The tangent vector at 0 of the curve $u(s) \otimes v(s) + h(s)$ in $\mathcal{S}^0$ is $a \otimes v + u \otimes b + k$. Differentiating the relation $h(s)u(s) = h^{\mathrm{T}}(s)v(s) = 0$ at $s = 0$ yields

$$ku = -ha \qquad \text{and} \qquad k^{\mathrm{T}}v = -h^{\mathrm{T}}b\,. \tag{10.3.1}$$

Taking $b = 0$, one sees that $k = -u \otimes (ha)$ satisfies (10.3.1) and so $a \otimes v - u \otimes (ha)$ is in the tangent space $T_{u\otimes v+h}\mathcal{S}^0$. Hence, $H^1_{u,v,h}$ is a subset of $T_{u\otimes v+h}\mathcal{S}^0$.

Considering $a = 0$ and checking that $k = (-h^{\mathrm{T}}b) \otimes v$ satisfies (10.3.1) yields the inclusion of $H^2_{u,v,h}$ in $T_{u\otimes v+h}\mathcal{S}^0$.

The orthogonality of $H^1_{u,v,h}$ and $H_{u,v}$ comes from the fact that for a in T_uS_{n-1} and h in $H_{u,v}$,

$$\langle a \otimes v - u \otimes (ha), h\rangle = \mathrm{tr}(av^{\mathrm{T}}h - u(ha)^{\mathrm{T}}h) = 0\,.$$

Similarly, one proves that $H^2_{u,v,h}$ is orthogonal to $H_{u,v}$.

As a consequence of Proposition 10.3.1, $\mathcal{S}^0$ is a manifold of dimension $n^2 - 1$ and $\dim H_{u,v} = (n-1)^2$. Thus,

$$\dim\big(H_{u,v} \oplus (H^1_{u,v,h} + H^2_{u,v,h})\big) = \dim \mathcal{S}^0$$

and we indeed found the whole tangent space to $\mathcal{S}^0$ — and not only a subspace. ∎

We can now construct explicitly an orthonormal basis for the tangent space in which we will express the second fundamental form of the immersion $\mathcal{S}^0 \subset \mathbb{R}^{n^2}$, and hence the curvature tensor of $\mathcal{S}^0$.

For this purpose, we denote by $e^u_1, \ldots, e^u_{n-1}$ an orthonormal basis of T_uS_{n-1}. Whenever h belongs to $H_{u,v}$, the vector u is in the kernel of $h^{\mathrm{T}}h$. Thus, if $\|h\| < 1$, the matrix $(\mathrm{Id} + h^{\mathrm{T}}h)^{-1/2}$ is well defined, and $\{\, u \,\}^{\perp}$ is an invariant subspace for this matrix. Consequently, the vectors

$$a_i = (\mathrm{Id} + h^{\mathrm{T}}h)^{-1/2}e^u_i \in T_uS_{n-1}\,, \qquad 1 \leqslant i \leqslant n-1\,,$$

are all in $T_u S_{n-1}$. They even span $T_u S_{n-1}$, because so do the e_i^u's and $\|h\| < 1$. The matrices

$$f_i = a_i \otimes v - u \otimes (ha_i)\,,\; 1 \leqslant i \leqslant n-1\,,$$

form an orthonormal basis of $H^1_{u,v,h}$ since an elementary calculation shows

$$\langle f_i, f_j \rangle = \langle a_i, (\mathrm{Id} + h^{\mathrm{T}} h) a_j \rangle = \delta_{i,j}\,.$$

To construct an orthonormal basis of the orthocomplement

$$K_{u,v,h} = (H^1_{u,v,h} + H^2_{u,v,h}) \ominus H^1_{u,v,h}\,,$$

notice that for h in $H_{u,v}$ with $\|h\| < 1$, the subspace $\{\, v \,\}^{\perp}$ is invariant under $(\mathrm{Id} + hh^{\mathrm{T}})^{1/2}$ and $(\mathrm{Id} - hh^{\mathrm{T}})^{-1}$. Thus,

$$b_j = (\mathrm{Id} - hh^{\mathrm{T}})^{-1} (\mathrm{Id} + hh^{\mathrm{T}})^{1/2} e_j^v \in T_v S_{n-1}\,.$$

For any b in $T_v S_{n-1}$, define

$$\begin{aligned} b^v &= b - 2 \sum_{1 \leqslant i \leqslant n-1} \langle b, ha_i \rangle h a_i \in \{\, v \,\}^{\perp}\,, \\ b^u &= h^{\mathrm{T}} b - 2 \sum_{1 \leqslant i \leqslant n-1} \langle b, ha_i \rangle a_i \in \{\, u \,\}^{\perp}\,. \end{aligned}$$

The vectors $b_j^u = (b_j)^u$ and $b_j^v = (b_j)^v$ are then defined, and so are the matrices

$$g_j = u \otimes b_j^v - b_j^u \otimes v \in \mathrm{M}(n, \mathbb{R})\,.$$

Using the bilinearity of the tensor product, we deduce that g_j belongs to $H^1_{u,v,h} + H^2_{u,v,h}$ since $u \otimes b_j - (h^{\mathrm{T}} b_j) \otimes v$ is in $H^2_{u,v,h}$ while $a_i \otimes v - u \otimes (ha_i)$ is in $H^1_{u,v,h}$.

10.3.3. PROPOSITION. *The matrices f_i, g_j, $1 \leqslant i, j \leqslant n-1$, form an orthonormal basis of $H^1_{u,v,h} + H^2_{u,v,h}$.*

Proof. It remains for us to prove that the g_j's are orthonormal, and that they are orthogonal to the f_i's. Since b_j^u is orthogonal to u and b_k^v to v,

$$\langle g_j, g_k \rangle = \langle b_j^v, b_k^v \rangle + \langle b_j^u, b_k^u \rangle\,.$$

Using the expression of b_j^u, b_k^u, b_j^v and b_k^v, we write

$$\langle g_j, g_k \rangle = \langle b_j, Q b_k \rangle\,,$$

where Q is the matrix

$$Q = hh^{\mathrm{T}} + \mathrm{Id} + 4 \sum_{1\leqslant i,l\leqslant n-1} ha_i a_l^{\mathrm{T}} h^{\mathrm{T}} \langle (\mathrm{Id} + h^{\mathrm{T}} h) a_i, a_l \rangle \\ - 8 \sum_{1\leqslant i\leqslant n-1} ha_i a_i^{\mathrm{T}} h^{\mathrm{T}} .$$

Since $\langle (\mathrm{Id} + h^{\mathrm{T}} h) a_i, a_l \rangle = \delta_{i,l}$, we have

$$Q = hh^{\mathrm{T}} + \mathrm{Id} - 4 \sum_{1\leqslant i\leqslant n-1} ha_i a_i^{\mathrm{T}} h^{\mathrm{T}} .$$

Notice that

$$\begin{aligned} \sum_{1\leqslant i\leqslant n-1} a_i a_i^{\mathrm{T}} &= \sum_{1\leqslant i\leqslant n-1} (\mathrm{Id} + h^{\mathrm{T}} h)^{-1/2} e_i^u e_i^{u\mathrm{T}} (\mathrm{Id} + h^{\mathrm{T}} h)^{-1/2} \\ &= (\mathrm{Id} + h^{\mathrm{T}} h)^{-1/2} \mathrm{Proj}_{u^\perp} (\mathrm{Id} + h^{\mathrm{T}} h)^{-1/2} . \end{aligned}$$

Since the image of h^{T} is orthogonal to u and $\{ u \}^\perp$ is invariant under $(\mathrm{Id} + h^{\mathrm{T}} h)^{-1/2}$, we obtain

$$Q = hh^{\mathrm{T}} + \mathrm{Id} - 4h(\mathrm{Id} + h^{\mathrm{T}} h)^{-1} h^{\mathrm{T}} .$$

This expression simplifies further since

$$\begin{aligned} &(\mathrm{Id} - hh^{\mathrm{T}})(\mathrm{Id} + hh^{\mathrm{T}})^{-1}(\mathrm{Id} - hh^{\mathrm{T}}) \\ &\quad = (\mathrm{Id} - hh^{\mathrm{T}}) \Big(\sum_{k\geqslant 0} (-1)^k (hh^{\mathrm{T}})^k - \sum_{k\geqslant 0} (-1)^k (hh^{\mathrm{T}})^{k+1} \Big) \\ &\quad = \sum_{k\geqslant 0} (-1)^k (hh^{\mathrm{T}})^k - \sum_{k\geqslant 0} (-1)^k (hh^{\mathrm{T}})^{k+1} \\ &\qquad\qquad - \sum_{k\geqslant 0} (-1)^k (hh^{\mathrm{T}})^{k+1} + \sum_{k\geqslant 0} (-1)^k (hh^{\mathrm{T}})^{k+2} \\ &\quad = \mathrm{Id} + hh^{\mathrm{T}} + 4 \sum_{k\geqslant 1} (-1)^k (hh^{\mathrm{T}})^k \\ &\quad = \mathrm{Id} + hh^{\mathrm{T}} - 4h(\mathrm{Id} + h^{\mathrm{T}} h)^{-1} h^{\mathrm{T}} \\ &\quad = Q \end{aligned}$$

Consequently, replacing b_j by its definition,

$$\begin{aligned} \langle g_j, g_k \rangle &= \langle (\mathrm{Id} - hh^{\mathrm{T}})^{-1} (\mathrm{Id} + hh^{\mathrm{T}})^{1/2} e_j^v , \\ &\qquad\qquad Q (\mathrm{Id} - hh^{\mathrm{T}})^{-1} (\mathrm{Id} + hh^{\mathrm{T}})^{1/2} e_k^v \rangle \\ &= \langle e_j^v, e_k^v \rangle = \delta_{j,k} . \end{aligned}$$

To conclude the proof, we calculate

$$\begin{aligned}\langle f_i, g_j\rangle &= \langle a_i \otimes v - u \otimes (ha_i)\,,\, u \otimes b_j^v - b_j^u \otimes v\rangle \\ &= -\langle a_i, b_j^u\rangle - \langle ha_i, b_j^v\rangle\,.\end{aligned}$$

Since

$$b_j^u + h^{\mathrm{T}} b_j^v = 2h^{\mathrm{T}} b_j - 2 \sum_{1\leqslant k\leqslant n-1} \langle b_j, ha_k\rangle (\mathrm{Id} + h^{\mathrm{T}} h) a_k\,,$$

we deduce that

$$\langle f_i, g_j\rangle = -2\langle a_i, h^{\mathrm{T}} b_j\rangle + 2 \sum_{1\leqslant k\leqslant n} \langle b_j, ha_k\rangle \delta_{i,k} = 0$$

as claimed. ■

Consider an orthonormal basis h_k, $1 \leqslant k \leqslant (n-1)^2$, of $H_{u,v}$. Furthermore, define the vectors

$$c_j = -\big[\mathrm{Id} - 2(\mathrm{Id} + h^{\mathrm{T}} h)^{-1}\big] h^{\mathrm{T}} b_j \in T_u S_{n-1}\,.$$

Quite remarkably, it is possible to explicitely calculate the curvature tensor of $\mathcal{S}^0$ through its second fundamental form.

10.3.4. THEOREM. *For u, v in S_{n-1}, for h in $H_{u,v}$ with $\|h\| < 1$, the second fundamental form of $\mathcal{S}$ at $u \otimes v + h$ in the orthogonal basis f_i, g_j, h_k, $1 \leqslant i,j \leqslant n-1$, $1 \leqslant k \leqslant (n-1)^2$, is given by the $(n^2-1)\times(n^2-1)$ matrix*

$$\Pi = \begin{array}{c} \\ {\scriptstyle n-1} \\ {\scriptstyle n-1} \\ {\scriptstyle n^2-2n+1} \end{array} \begin{array}{c} \begin{array}{ccc} {\scriptstyle n-1} & \qquad\qquad {\scriptstyle n-1} \qquad\qquad & {\scriptstyle n^2-2n+1} \end{array} \\ \begin{pmatrix} \langle a_i, a_j\rangle & \langle c_i, a_j\rangle & 0 \\ \langle a_i, c_j\rangle & \langle c_i, c_j\rangle + \langle b_i, (\mathrm{Id} - hh^{\mathrm{T}}) b_j\rangle & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array}$$

Proof. Let $N = u\otimes v$ be the outward unit normal to $\mathcal{S}^0$ at $u\otimes v + h$ — see Proposition 10.3.2. We will denote by ∇ the covariant derivative on $\mathcal{S}^0$; that is, for a vector field X defined on $\mathcal{S}^0$, for u a tangent vector vector field and p a point on $\mathcal{S}^0$,

$$\nabla_u X(p) = \mathrm{Proj}_{T_p\mathcal{S}^0} \mathrm{D}X(p)\cdot u\,.$$

The calculations made in the proof of Proposition 10.3.2 show that f_i is the tangent vector at 0 of a curve $s \mapsto u(s) \otimes v + h(s)$ with $u(0) = u$, $h(0) = h$ and $u'(0) = a_i$ and $h'(0) = -u \otimes h a_i$. Consequently $\nabla_{f_i} N = a_i \otimes v$. Moreover,

$$\begin{aligned}\nabla_{g_j} N &= \nabla_{u\otimes b_j - (h^{\mathrm{T}} b_j \otimes v)} N - 2 \sum_{1 \leqslant i \leqslant n-1} \langle b_j, h a_i \rangle \nabla_{u \otimes (h a_i) - a_i \otimes v} N \\ &= u \otimes b_j + 2 \sum_{1 \leqslant i \leqslant n-1} \langle b_j, h a_i \rangle a_i \otimes v \,.\end{aligned}$$

Moreover, since N is constant along $h \in H_{u,v} \mapsto u \otimes v + h$, $\|h\| < 1$, we have $\nabla_{h_k} N = 0$. A routine calculation gives the first entries of the matrix, namely

$$\langle \nabla_{f_i} N, f_j \rangle = \langle a_i, a_j \rangle \,.$$

Next, we have

$$\begin{aligned}\langle \nabla_{g_i} N, f_j \rangle &= -\langle h^{\mathrm{T}} b_i, a_j \rangle + 2 \sum_{1 \leqslant k \leqslant n-1} \langle h^{\mathrm{T}} b_i, a_k \rangle \langle a_k, a_j \rangle \\ &= \langle a_j, -h^{\mathrm{T}} b_i + 2 \sum_{1 \leqslant k \leqslant n-1} a_k a_k^{\mathrm{T}} h^{\mathrm{T}} b_i \rangle \,.\end{aligned}$$

Since the image of h^{T} is orthogonal to u,

$$\sum_{1 \leqslant k \leqslant n-1} a_k a_k^{\mathrm{T}} h^{\mathrm{T}} = (\mathrm{Id} + h h^{\mathrm{T}})^{-1} h^{\mathrm{T}} \,.$$

This gives the entries in $\langle a_i, c_j \rangle$.

Finally, we calculate

$$\begin{aligned}\langle \nabla_{g_i} N, g_j \rangle &= \big\langle u \otimes b_i + 2 \sum_{1 \leqslant k \leqslant n-1} \langle b_i, h a_k \rangle a_k \otimes v \,,\, u \otimes b_j^v - b_j^u \otimes v \big\rangle \\ &= \langle b_i, b_j^v \rangle - 2 \sum_{1 \leqslant k \leqslant n-1} \langle b_i, h a_k \rangle \langle a_k, b_j^u \rangle \\ &= \langle b_i, Q b_j \rangle \,,\end{aligned}$$

where the matrix Q is

$$Q = \mathrm{Id} - 4 \sum_{1 \leqslant l \leqslant n-1} h a_l a_l^{\mathrm{T}} h^{\mathrm{T}} + 4 \sum_{1 \leqslant k,l \leqslant n-1} h a_k a_l^{\mathrm{T}} h^{\mathrm{T}} \langle a_k, a_l \rangle \,.$$

Again, since the image of h^{T} is orthogonal to u,

$$h \sum_{1 \leqslant l \leqslant n-1} a_l a_l^{\mathrm{T}} h^{\mathrm{T}} = h(\mathrm{Id} + h h^{\mathrm{T}})^{-1} h^{\mathrm{T}} \,.$$

Moreover,

$$\begin{aligned} h &\sum_{1\leqslant k,l\leqslant n-1} a_k a_l^{\mathrm{T}} \langle a_k, a_l\rangle h^{\mathrm{T}} \\ &= h(\mathrm{Id}+hh^{\mathrm{T}})^{-1/2} \sum_{1\leqslant k,l\leqslant n-1} e_k e_l^{\mathrm{T}} \langle e_k, (\mathrm{Id}+h^{\mathrm{T}}h)^{-1} e_l\rangle \times \\ &\qquad\qquad (\mathrm{Id}+hh^{\mathrm{T}})^{-1/2} h^{\mathrm{T}} \\ &= h(\mathrm{Id}+hh^{\mathrm{T}})^{-2} h^{\mathrm{T}} . \end{aligned}$$

Consequently,

$$\begin{aligned} Q &= \mathrm{Id} - 4h(\mathrm{Id}+hh^{\mathrm{T}})^{-1}h^{\mathrm{T}} + 4h(\mathrm{Id}+hh^{\mathrm{T}})^{-2}h^{\mathrm{T}} \\ &= \mathrm{Id} - hh^{\mathrm{T}} + h\big(\mathrm{Id} - 2(\mathrm{Id}+hh^{\mathrm{T}})^{-1}\big)^2 h^{\mathrm{T}} . \end{aligned}$$

This gives

$$\langle \nabla_{g_i} N, g_j\rangle = \langle b_i, (\mathrm{Id}-hh^{\mathrm{T}})b_j\rangle + \langle c_i, c_j\rangle$$

as claimed. ∎

It follows from Theorem 10.3.4 and elementary results on immersions that the Riemannian curvature tensor R of $\mathcal{S}^0$ can be calculated explicitly in the basis f_i, g_j, h_k. It is convenient to define $f_{n-1+i} = g_i$ for $i = 1, \ldots, n-1$ and $f_{2(n-1)+i} = h_i$ for $i = 1, \ldots, (n-1)^2$. If X, Y are two elements of $T_{u\otimes v+h}\mathcal{S}^0$, then

$$\langle R(f_i, f_j)X, Y\rangle = X^{\mathrm{T}}\big(\nabla_{f_i}N(\nabla_{f_j}N)^{\mathrm{T}} - \nabla_{f_j}N(\nabla_{f_i}N)^{\mathrm{T}}\big)Y .$$

Thus, $R(f_i, f_j)$ is the compression to the tangent space of the matrix $\nabla_{f_i}N(\nabla_{f_j}N)^{\mathrm{T}} - \nabla_{f_j}N(\nabla_{f_i}N)^{\mathrm{T}}$.

As a byproduct of the work done, we can prove that there are no flat and nontrivial convex subsets in $\mathcal{S}^0$ besides the unit balls of the fibers $H_{u,v}$. This result will not be used in the sequel, but brings more intuition on the shape of the sphere $\mathcal{S}$. It is enough to show that the $2(n-1)\times 2(n-1)$ upper left corner submatrix of Π has no zero eigenvalue. Since $\|h\| < 1$ on our parameterization of $\mathcal{S}^0$, this follows from the next result.

10.3.5. PROPOSITION. *The following equality holds,*

$$\begin{aligned} \det &\begin{pmatrix} \langle a_i, a_j\rangle_{i,j} & \langle a_i, c_j\rangle_{i,j} \\ \langle a_i, c_j\rangle_{i,j} & \langle c_i, c_j\rangle_{i,j} + \langle b_i, (\mathrm{Id}-hh^{\mathrm{T}})b_j\rangle_{i,j} \end{pmatrix} \\ &\qquad = \det(\mathrm{Id}+hh^{\mathrm{T}})\det(\mathrm{Id}+h^{\mathrm{T}}h)\det(\mathrm{Id}-hh^{\mathrm{T}}) . \end{aligned}$$

Proof. We first calculate a subdeterminant of the given one. Going back to the definition of the a_i's and using that u is an eigenvector of $(\mathrm{Id}+h^{\mathrm{T}}h)$ associated to the eigenvalue 1,

$$\begin{aligned}\det(\langle a_i, a_j\rangle)_{i,j} &= \det\bigl(e_i^u, (\mathrm{Id}+h^{\mathrm{T}}h)e_j^u\rangle\bigr)_{1\leqslant i,j\leqslant n-1}\\ &= \det(\mathrm{Id}+h^{\mathrm{T}}h)\,.\end{aligned}$$

Furthermore, since v is an eigenvalue of $(\mathrm{Id}+h^{\mathrm{T}}h)^{1/2}$ and $(\mathrm{Id}-hh^{\mathrm{T}})^{-1}$ associated with the eigenvalue 1 — this can be seen by series expanding and using the fact that $h^{\mathrm{T}}v=0$ —

$$\begin{aligned}&\det\bigl(\langle b_i, (\mathrm{Id}-hh^{\mathrm{T}})b_j\rangle\bigr)_{i,j}\\ &\quad= \det\bigl(\langle e_i^v, (\mathrm{Id}+hh^{\mathrm{T}})^{1/2}(\mathrm{Id}-hh^{\mathrm{T}})^{-1}(\mathrm{Id}+hh^{\mathrm{T}})^{1/2}e_j^v\rangle_{i,i}\bigr)\\ &\quad= \det(\mathrm{Id}+hh^{\mathrm{T}})\det(\mathrm{Id}-hh^{\mathrm{T}})\,.\end{aligned}$$

To conclude the proof, we use the following claim, with $d_i=(\mathrm{Id}-hh^{\mathrm{T}})^{1/2}b_i$.

Claim. *Let a_i, c_i, d_i, $1\leqslant i\leqslant n-1$, be $3(n-1)$ vectors in $\mathbb{R}^n$. Consider the $n\times(n-1)$-matrices $a=(a_1,\dots,a_{n-1})$, $c=(c_1,\dots,c_{n-1})$ and $d=(d_1,\dots,d_{n-1})$. If the image of c is contained in the image of a, then*

$$\det\left(\begin{pmatrix} a^{\mathrm{T}}\\ c^{\mathrm{T}}\end{pmatrix}\begin{pmatrix} a & c\end{pmatrix}+\begin{pmatrix}0 & 0\\ 0 & d^{\mathrm{T}}d\end{pmatrix}\right)=\det(a^{\mathrm{T}}a)\det(d^{\mathrm{T}}d)\,.$$

To prove the claim, let P be an orthogonal matrix and D be a diagonal one such that $d^{\mathrm{T}}d=PDP^{\mathrm{T}}$. Writing M for the matrix whose determinant we want to calculate, we have

$$\begin{aligned}\det M &= \det\left(\begin{pmatrix}\mathrm{Id} & 0\\ 0 & P\end{pmatrix}\begin{pmatrix} a^{\mathrm{T}}\\ c^{\mathrm{T}}\end{pmatrix}\begin{pmatrix} a & c\end{pmatrix}\begin{pmatrix}\mathrm{Id} & 0\\ 0 & P^{\mathrm{T}}\end{pmatrix}+\begin{pmatrix}0 & 0\\ 0 & D\end{pmatrix}\right)\\ &= \det\left(V+\begin{pmatrix}0 & 0\\ 0 & D\end{pmatrix}\right).\end{aligned}$$

The proof then goes by induction on the dimension of D, noticing that for any real number δ,

$$\det\left(V+\begin{pmatrix}0 & 0\\ 0 & \delta\end{pmatrix}\right)=\delta\det\widehat{V}_{m,m}+\det V$$

where $\widehat{V}_{m,m}$ is the $(m-1)\times(m-1)$ left upper corner of V and m the dimension of V. Consequently, we just need to prove that $\det V = 0$. This is clear since the condition $\mathrm{Im}c \subset \mathrm{Im}a$ implies that the rank of the matrix $(\, a \quad c\,)$ is the dimension of the image of a, and hence the rank of V is at most $\dim \mathrm{Im}a$. Consequently, we have

$$\det\left(V + \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \right) = \det{(a^{\mathrm{T}}a)} \det D\,.$$

This proves the claim and concludes the proof of Proposition 10.3.5.■

As a consequence of Proposition 10.3.5, the Gauss-Kronecker curvature of the nonflat part of $\mathcal{S}^0$ at $u\otimes v + h$ is given by $\det(\mathrm{Id} + h^{\mathrm{T}}h)\det(\mathrm{Id}+hh^{\mathrm{T}})\det(\mathrm{Id}-hh^{\mathrm{T}})$. Other curvatures can be calculated as well, leading to more or less interesting formulas.

10.4. Norms of random matrices.

Let us again consider a random matrix $X = (X_{i,j})_{1\leqslant i,j\leqslant n}$ with independent and identically distributed coefficients. Its (operator) norm is

$$\|X\| = \sup\{\, |Xu| : |u| = 1\,,\ u\in\mathbb{R}^n\,\}\,.$$

In this section, we will obtain estimates for the tail probability $P\{\,\|X\|\geqslant t\,\}$, assuming that the $X_{i,j}$'s are either symmetric Weibull or Student like distributed.

In theory, we just need to apply the results of chapter 9. Indeed, $\mathbb{R}^n$ being reflexive, $\|X\| = \sup\{\,\langle Xu,v\rangle : u,v\in S_{n-1}\,\}$. Since $\langle Xu,v\rangle_{\mathbb{R}^n} = \langle X, u\otimes v\rangle_{\mathbb{R}^{n^2}}$, we see that $\|X\|$ is the supremum of the linear form X acting on the submanifold $S_{n-1}\otimes S_{n-1}$ of $\mathbb{R}^{n^2}$.

However, a direct application of the results of chapter 9 in the case of light tails is not that easy. We will proceed by using both chapter 7 and ideas from sections 9.1 and 9.2 as well.

Our first result is for light tails.

10.4.1. THEOREM. *Let $X = (X_{i,j})_{1\leqslant i,j\leqslant n}$ be a random matrix with independent and identically distributed coefficients, all having the symmetric Weibull-like density*

$$w_\alpha(x) = \frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)}\exp\Big(\frac{-|x|^\alpha}{\alpha}\Big)\,,\qquad x\in\mathbb{R}\,.$$

(i) If $\alpha = 2$, then

$$P\{\, \|X\| \geqslant t \,\} \sim \frac{\sqrt{2\pi}}{2^{n-1}\Gamma(n/2)^2} e^{-t^2/2} t^{2(n-1)-1} \qquad \text{as } t \to \infty\,.$$

(ii) If $\alpha > 2$, then

$$P\{\, \|X\| \geqslant t \,\} \sim \left(\frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)}\right)^{n^2} \times \frac{(2\pi)^{(n^2-1)/2} 2^{2n-1}}{n^{(2-\alpha)(n^2+1)/2}(\alpha-2)^{n-1}(\alpha-1)^{(n-1)^2/2}} \frac{e^{-t^\alpha/(\alpha n^{\alpha-2})}}{t^{\frac{\alpha}{2}(n^2+1)-n^2}}$$

as t tends to infinity.

Proof. Define

$$A_t = \left\{\, x = (x_{i,j})_{1 \leqslant i,j \leqslant n} \in \mathrm{M}(n, \mathbb{R}) : \|x\| \geqslant t \,\right\} = tA_1\,.$$

We need to evaluate

$$\left(\frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)}\right)^{n^2} \int_{tA_1} \exp\left(-\frac{1}{\alpha}\sum_{1 \leqslant i,j \leqslant n} |x_{i,j}|^\alpha\right) \prod_{1 \leqslant i,j \leqslant n} \mathrm{d}x_{i,j}\,.$$

Defines the α-homogeneous function

$$I(x) = \frac{|x|_\alpha^\alpha}{\alpha} = \frac{1}{\alpha}\sum_{1 \leqslant i,j \leqslant n} |x_{i,j}|^\alpha\,.$$

We can apply Theorem 7.1. The first step is to calculate $I(A_1)$ and $\mathcal{D}_{A_1}$. To do this, we need a description of the boundary

$$\partial A_1 = \left\{\, x \in \mathrm{M}(n, \mathbb{R}) : \|x\| = 1 \,\right\},$$

that is of the sphere of radius 1 in the space of matrices endowed with the operator norm. This is provided by Proposition 10.3.1. Let us simply recall here that the matrices of norm 1 coincide with all matrices of the form $u \otimes v + h$, where u, v belong to the sphere S_{n-1} and h is an $n \times n$ matrix of operator norm less than 1, satisfying $hu = h^T v = 0$. This allows us to find $\mathcal{D}_{A_1}$.

10.4.2. LEMMA. *The function I is minimum over ∂A_1 exactly at matrices of the form $u \otimes v$ with*
(i) $u, v \in \left\{\, n^{-1/2}(\epsilon_1, \ldots, \epsilon_n) : \epsilon_i \in \{\, -1, 1 \,\},\ 1 \leqslant i \leqslant n \,\right\}$ if $\alpha > 2$,

(ii) $u, v \in S_{n-1}$ *if* $\alpha = 2$,
(iii) $u, v \in \{\, \epsilon e_i : \epsilon \in \{-1, 1\}\,,\ 1 \leqslant i \leqslant n \,\}$, *if* $\alpha < 2$.

Proof. As we already mentioned, $\|x\|$ is the supremum of the linear form $x \in \mathbb{R}^{n^2}$ acting on $S_{n-1} \otimes S_{n-1}$. Proposition 9.1.7 and convexity of A_1^{c} implies $I(\partial A_1) = I_\bullet(S_{n-1} \otimes S_{n-1})$. Moreover, Lemma 9.1.9 asserts that $I_\bullet(x) = 1/(\alpha |x|_\beta^\alpha)$ where $\alpha^{-1} + \beta^{-1} = 1$. We can first calculate the points in $S_{n-1} \otimes S_{n-1}$ which minimize $I_\bullet$. This is rather easy since

$$\alpha I_\bullet(u \otimes v) = \Big(\sum_{1 \leqslant i,j \leqslant n} |v_i u_j|^\beta \Big)^{-\alpha/\beta} = |v|_\beta^{-\alpha} |u|_\beta^{-\alpha} \,.$$

Thus, we need to locate the maxima of $|u|_\beta$ on S_{n-1}.

If α is larger than 2, then β is smaller than 2. Therefore,

$$\Big(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} |u_i|^\beta \Big)^{1/\beta} \leqslant \Big(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} u_i^2 \Big)^{1/2} = \frac{1}{\sqrt{n}} \,,$$

with equality if and only if $|u_i| = 1/\sqrt{n}$ for all $i = 1, 2, \ldots, n$. Consequently,

$$\sup \{\, |u|_\beta : u \in S_{n-1} \,\} = n^{\frac{1}{\beta} - \frac{1}{2}} \,, \qquad \text{if } \alpha > 2 \,.$$

If $\alpha = 2$, then $\beta = 2$, and $|u|_\beta = 1$ over all S_{n-1}.

Finally, if α is smaller than 2, then β is larger than 2. A unit vector u has all its components u_i between -1 and 1. Therefore,

$$|u|_\beta = \Big(\sum_{1 \leqslant i \leqslant n} |u_i|^\beta \Big)^{1/\beta} \leqslant \Big(\sum_{1 \leqslant i \leqslant n} |u_i|^2 \Big)^{1/\beta} = 1 \,,$$

with equality if and only if one — and only one — of the $|u_i|$'s is 1.

Consequently,

$$\alpha I_\bullet(u \otimes v) \geqslant \begin{cases} n^{2-\alpha} & \text{if } \alpha > 2, \\ 1 & \text{if } \alpha \leqslant 2, \end{cases}$$

with equality for $(u, v) = (u_*, v_*)$ with (u_*, v_*) exactly in the following sets,

$$\begin{cases} |u_{*,i}| = |v_{*,i}| = 1/\sqrt{n} & i = 1, 2, \ldots, n, \text{ if } \alpha > 2, \\ u_*, v_* \in S_{n-1} & \text{if } \alpha = 2, \\ u_*, v_* \in \{\, \epsilon e_i : \epsilon \in \{-1, 1\}\,,\ 1 \leqslant i \leqslant n \,\} & \text{if } \alpha < 2. \end{cases}$$

For u and v in S_{n-1}, set

$$H_{u,v} = \left\{ h \in \mathrm{M}(n, \mathbb{R}) : hu = h^T v = 0 \right\}.$$

For h in $H_{u,v}$, we have

$$\langle u \otimes v, u \otimes v + h \rangle = \mathrm{tr}(uv^T vu^T + uv^T h) = 1\,.$$

Consequently,

$$\begin{aligned} I_\bullet(u \otimes v) &= \inf \left\{ I(x) : \langle x, u \otimes v \rangle = 1 \right\} \\ &\leqslant \inf \left\{ I(u \otimes v + h) : h \in H_{u,v}\,,\ \|h\| \leqslant 1 \right\}. \end{aligned}$$

The inequality

$$I_\bullet(u_* \otimes v_*) \leqslant \inf \left\{ I(x) : x \in \partial A_1 \right\}.$$

follows. Observe that $I_\bullet(u_* \otimes v_*) = I(u_* \otimes v_*)$ for any $\alpha \geqslant 1$. Since the function $h \in H_{u,v} \mapsto I(u \otimes v + h)$ is convex, as a restriction of a convex function to a convex set, the infimum of I over ∂A_1 is achieved only at points $x = u_* \otimes v_*$. On those points $I_\bullet$ and I coincide and this concludes the proof. ∎

It is interesting to realize that the proof of Lemma 10.4.2 relies on the fact that $I_\bullet$ and I coincide on the matrices $u_* \otimes v_*$. Geometrically, the matrices $u \otimes v + h$, $h \in H_{u,v}$ with $\|h\| \leqslant 1$ forms a truncated cylinder with base $S_{n-1} \otimes S_{n-1}$. What makes the proof work is that S_{n-1} is the polar reciprocal of its convex hull; a very special property of the sphere!

Let us now calculate all the terms that come from applying Theorem 7.1. We will then justify that we can indeed apply this theorem in verifying that its assumptions hold.

Let us first consider the case $\alpha > 2$. From Lemma 10.4.2, we deduce

$$I(A_1) = n^{2-\alpha}/\alpha\,.$$

The rescaled dominating manifold

$$\mathcal{D}_{A_1} = \Big\{ u \otimes v : u, v \text{ of the form } \frac{1}{\sqrt{n}}(\epsilon_1, \ldots, \epsilon_n) : \\ \epsilon_i \in \{ -1, 1 \}\,,\ 1 \leqslant i \leqslant n \Big\}.$$

is of dimension $k=0$. Its Riemannian volume is the counting measure

$$\mathcal{M}_{\mathcal{D}_{A_1}} = \frac{1}{2}\sum_{1\leqslant i,j\leqslant n}\sum_{\epsilon_i,\eta_j\in\{-1,1\}} \delta_{\frac{1}{n}(\epsilon_1,\ldots,\epsilon_n)\otimes(\eta_1,\ldots,\eta_n)} = \sum \delta_{\epsilon\otimes\eta/n}$$

where the last sum is over all distinct matrices $(\eta_i\epsilon_j/n)_{1\leqslant i,j\leqslant n}$ with coefficients in $\{\,-1/n, 1/n\,\}$. Note that $\epsilon\otimes\eta$ and $(-\epsilon)\otimes(-\eta)$ are equal, thus not distinct.

For x in $\mathcal{D}_{A_1}$, we have

$$|\mathrm{D}I(x)|^2 = \sum_{1\leqslant i,j\leqslant n} |x_{i,j}|^{2(\alpha-1)} = \frac{n^2}{n^{2(\alpha-1)}} = n^{4-2\alpha}\,.$$

We then need to calculate the curvature term $\det G_{A_1}$, and hence the fundamental form $\Pi_{\Lambda_{I(A_1)},u_*\otimes v_*}$ and $\Pi_{\partial A_1,u_*\otimes v_*}$ for $u_*\otimes v_*$ in $\mathcal{D}_{A_1}$.

From Theorem 10.3.4 with $h=0$, we deduce that

$$\Pi_{\partial A_1} = \begin{pmatrix} \mathrm{Id}_{2(n-1)} & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{M}(\mathrm{n}^2-1, \mathbb{R}\,)\,.$$

On the other hand, the second fundamental form of $\Lambda_{I(A_1)}$ at x is the restriction to the tangent space $T_x\Lambda_{I(A_1)}$ of

$$\frac{\mathrm{D}^2 I(x)}{|\mathrm{D}I(x)|} = \Big(\sum_{1\leqslant i,j\leqslant n} |x_{i,j}|^{2(\alpha-1)}\Big)^{-1/2} \mathrm{diag}\big(|x_{i,j}|^{\alpha-2}\big)_{1\leqslant i,j\leqslant n}(\alpha-1)\,.$$

Thus, if x is in $\mathcal{D}_{A_1}$, we have $|x_{i,j}|=1/n$ and

$$\frac{\mathrm{D}^2 I(x)}{|\mathrm{D}I(x)|} = \frac{1}{n^{2-\alpha}}\frac{\alpha-1}{n^{\alpha-2}}\mathrm{Id}_{n^2} = (\alpha-1)\mathrm{Id}_{n^2}\,.$$

Hence, for x in $\mathcal{D}_{A_1}$,

$$\begin{aligned}\Pi_{\Lambda_{I(A_1)},x} - \Pi_{\partial A_1,x} &= (\alpha-1)\mathrm{Id}_{n^2-1} - \begin{pmatrix} \mathrm{Id}_{2(n-1)} & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} (\alpha-2)\mathrm{Id}_{2(n-1)} & 0 \\ 0 & (\alpha-1)\mathrm{Id}_{n^2-2n+1} \end{pmatrix}\,.\end{aligned}$$

Therefore, on $\mathcal{D}_{A_1}$,

$$\det G_{A_1}(x) = (\alpha-2)^{2(n-1)}(\alpha-1)^{n^2-2n+1}\,.$$

We obtain the constant c_1 in Theorem 7.1,

$$\begin{aligned} c_1 &= (2\pi)^{(n^2-1)/2}\Big(\frac{1}{n^{2-\alpha}}\Big)^{(n^2+1)/2}\frac{1}{(\alpha-2)^{n-1}(\alpha-1)^{(n-1)^2/2}}\sharp\mathcal{D}_{A_1} \\ &= \frac{(2\pi)^{(n^2-1)/2}2^{2n-1}}{n^{(2-\alpha)(n^2+1)/2}(\alpha-2)^{n-1}(\alpha-1)^{(n-1)^2/2}} \end{aligned}$$

Putting all the pieces together,

$$P(A_t) \sim \Big(\frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)}\Big)^{n^2}\frac{e^{-n^{2-\alpha}t^\alpha/\alpha}}{t^{(\alpha-2)\frac{n^2}{2}+\frac{\alpha}{2}}}\times \\ \frac{(2\pi)^{(n^2-1)/2}2^{2n-1}}{n^{(2-\alpha)(n^2+1)/2}(\alpha-2)^{n-1}(\alpha-1)^{(n-1)^2/2}}$$

as t tends to infinity, which is the result.

Let us now turn to the case $\alpha = 2$. From Lemma 10.4.2, we conclude

$$I(A_1) = 1/2\,.$$

The dominating manifold

$$\mathcal{D}_{A_1} = S_{n-1}\otimes S_{n-1}$$

is now of dimension $k = 2(n-1)$. Since $S_{n-1}\times S_{n-1}$ is a double covering of $\mathcal{D}_{A_1}$, the Riemannian measure on $\mathcal{D}_{A_1}$ is half the product measure on the product of 2 spheres S_{n-1}, each having the Riemannian measure obtained from the Lebesgue measure on $\mathbb{R}^n$.

On $\mathcal{D}_{A_1}$, we also have

$$|\mathrm{D}I(x)| = \Big(\sum_{1\leqslant i,j\leqslant n}|u_iv_j|^2\Big)^{1/2} = |u||v| = 1\,.$$

The computation of the curvature term $\det G_{A_1}$ goes the same way as for $\alpha > 2$. Namely, we still have

$$\Pi_{\Lambda_{I(A_1)}} = \mathrm{Id}_{n^2-1} \qquad \text{and} \qquad \Pi_{\partial A_1} = \begin{pmatrix} \mathrm{Id}_{2(n-1)} & 0 \\ 0 & 0 \end{pmatrix}.$$

In particular,

$$\Pi_{\Lambda_{I(A_1)}} - \Pi_{\partial A_1} = \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{Id}_{n^2-1-2(n-1)} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{Id}_{(n-1)^2} \end{pmatrix}.$$

It follows that

$$G_{A_1} = \mathrm{Id}_{(n-1)^2}$$

is of determinant 1. Therefore, with the notation of Theorem 7.1,

$$c_1 = (2\pi)^{(n^2-2(n-1)-1)/2} \frac{\mathrm{Vol}(S_{n-1})^2}{2} = (2\pi)^{(n-1)^2/2} \frac{\mathrm{Vol}(S_{n-1})^2}{2}\,.$$

Again, taking all the above estimates into account, we obtain

$$P\{\,\|X\| \geqslant t\,\} \sim \frac{1}{2}(2\pi)^{((n-1)^2-n^2)/2} \mathrm{Vol}(S_{n-1})^2 e^{-t^2/2} t^{2(n-1)-1}$$

as t tends to infinity. This is the result since S_{n-1} has volume $2\pi^{n/2}/\Gamma(n/2)$.

It remains to check the assumptions of Theorem 7.1. We already checked (7.3) and (7.4). Assumption (7.5) is clear as well since the curvature of ∂A_1 is bounded. ∎

We can now apply Theorem 7.5 to obtain the following result on conditional distributions. It is worth knowing that $S_1 \times S_1/\{\,-\mathrm{Id}, \mathrm{Id}\,\}$ is the usual Klein bottle. Hence, $S_{n-1} \otimes S_{n-1} \equiv S_{n-1} \times S_{n-1}/\{\,-\mathrm{Id}, \mathrm{Id}\,\}$ is a $2(n-1)$-dimensional Klein bottle.

10.4.3. PROPOSITION. *Let $X = (X_{i,j})_{1 \leqslant i,j \leqslant n}$ be a random matrix with independent and identically distributed coefficients, all having the density*

$$w_\alpha(s) = \frac{\alpha^{1-(1/\alpha)}}{2\Gamma(1/\alpha)} e^{-|s|^\alpha/\alpha}\,, \qquad s \in \mathbb{R}\,.$$

The conditional distribution of X/t given $\|X\| \geqslant t$ converges weakly to a uniform distribution over*
(i) the Klein bottle $S_{n-1} \otimes S_{n-1}$ if $\alpha = 2$,
(ii) the 2^{2n-1} matrices of the form $u_* \otimes v_*$ for $|u_{*,i}| = |v_{*,i}| = 1/\sqrt{n}$, $1 \leqslant i \leqslant n$, if $\alpha > 2$.

Proof. It follows from the calculation of $\mathcal{D}_{A_1}$ made in the proof of Theorem 10.4.1 and Theorem 7.5. ∎

Let us now turn to the problem of estimating the tail probability of $\|X\|$ when the coefficients $X_{i,j}$ of the random matrix X are independent and identically distributed with a Student-like distribution. Given the work done in the previous sections, this turns to be an easy problem.

10.4.4. THEOREM. *Let $X = (X_{i,j})_{1 \leqslant i,j \leqslant n}$ be a random matrix with independent coefficients, all having a Student-like distribution with parameter α. Then,*

$$P\{ \|X\| \geqslant t \} \sim \frac{2n^2 K_{s,\alpha} \alpha^{(\alpha-1)/2}}{t^\alpha} \qquad \text{as } t \to \infty\,.$$

Proof. Notice that the set

$$A_t = \{ x \in \mathrm{M}(n, \mathbb{R}) : \|x\| \geqslant t \} = tA_1$$

is the complement of a convex set, namely the ball of radius t centered at the origin in $\mathrm{M}(n, \mathbb{R})$ endowed with the operator norm. In $\mathrm{M}(n, \mathbb{R}) \equiv \mathbb{R}^{n^2}$, the axial points of this convex set are all the matrices $\epsilon E^{i,j}$, $\epsilon \in \{ -1, 1 \}$, $1 \leqslant i, j \leqslant n$, which are of Euclidean norm 1. There are $2n^2$ such matrices. Apply Theorem 9.3.1 to obtain the result. ∎

With no extra effort, we can also obtain the following result on conditional distribution.

10.4.5. PROPOSITION. *Let $X = (X_{i,j})_{1 \leqslant i,j \leqslant n}$ be a random matrix with independent coefficients, all having a Student-like distribution with parameter α. The distribution of X/t given $\|X\| \geqslant t$ converges weakly* to the uniform mixture of the distributions of the matrices $\epsilon Z E^{i,j}$ with ϵ in $\{ -1, 1 \}$ and Z having a Pareto distribution,*

$$P\{ Z \geqslant 1 + \lambda \} = \frac{1}{(1+\lambda)^\alpha}\,, \qquad \lambda \geqslant 0\,.$$

Proof. Apply Corollary 9.3.4. The axial points of A_1^{c} are the matrices $\epsilon E^{i,j}$, where ϵ is in $\{ -1, 1 \}$. Those matrices are of unit Euclidean norm in $\mathbb{R}^{n^2}$. ∎

Notes

The theory of random matrices has been evolving quite fast lately. Motivated by applications in physics and in operator algebras, important progress has been made on the asymptotic theory as the size of the matrix goes to infinity. Our fixed size viewpoint is quite different. Amazingly clever explicit calculations have been made in the Gaussian

cases and some of its variations. A classical reference is Mehta (1991). Another aspect driven by statistics concerns the Wishart distribution — see Johnson and Kotz (1972).

I believe Lemma 10.2.3 is not new, but I have not found a reference for it. It is very similar to Theorem 2.1 of Rosiński and Woyczyński (1987), as well as its proof. If we assume that the X_i's are symmetric, then Lemma 10.2.3 can be deduced from Rosiński and Woyczyński (1987) in conditioning on the signs of the X_i's. But here, we assume only asymptotic symmetry of the tail. Therefore, the signs of X_i's given $|X_i|$ large is only asymptotically distributed uniformly over $\{-1,+1\}$.

Theorem 10.1.4 involves the volume of $\mathrm{SO}(n,\mathbb{R})$. It seems to be calculated in Marinov (1980). But, by ignorance, I have not been able to follow his proof. I don't know if Marinov's results give the volume of $\mathrm{SO}(n,\mathbb{R})$ embedded in $\mathbb{R}^{n^2}$ of if they give it up to a proportionality constant.

11. Finite sample results for autoregressive processes

Autoregressive models are among the simplest and most widely used models in statistical analysis of time series. Their classical theory deals mainly with their asymptotic behavior over a very large time period. In this chapter, we will see that these models are in fact much more subtle than usually believed. Our study will build upon results of Chapter 8. Our results will not exhaust the topic by any mean; they should be considered as an incentive for further study.

11.1. Background on autoregressive processes.

In order to describe the processes we are interested in, let us introduce the backward shift B on vectors. For a vector $u = (u_1, \ldots, u_n)$ in $\mathbb{R}^n$, we write $Bu = (0, u_1, \ldots, u_n)$. Let ϵ be a mean zero random vector in $\mathbb{R}^n$, with independent and identically distributed components. We say that the vector X in $\mathbb{R}^n$ is an autoregressive process of order p with innovation ϵ if for some $\theta = (\theta_1, \ldots, \theta_p)$ in $\mathbb{R}^p$, with θ_p not null, it satisfies the equation

$$X = \sum_{1 \leqslant i \leqslant p} \theta_i B^i X + \epsilon \,. \tag{11.1.1}$$

In a perhaps more explicit form, that means

$$\begin{aligned}
X_1 &= \epsilon_1 \\
X_2 &= \theta_1 X_1 + \epsilon_2 \\
&\vdots \\
X_{p+1} &= \theta_1 X_p + \theta_2 X_{p-1} + \cdots + \theta_p X_1 + \epsilon_{p+1} \\
X_{p+2} &= \theta_1 X_{p+1} + \theta_2 X_p + \cdots + \theta_p X_2 + \epsilon_{p+2} \\
&\vdots \\
X_n &= \theta_1 X_{n-1} + \theta_2 X_{n-2} + \cdots + \theta_p X_{n-p} + \epsilon_n \,.
\end{aligned}$$

For statisticians, the main questions are on estimation and tests procedures for such models. This means that one observes the vector X, and knows that it is of the form (11.1.1) with the ϵ_i's independent

and identically distributed. The goal is then to estimate θ, that is to guess its value based on the knowledge of X; or to perform tests on θ, that is to check if some assumption on θ is compatible with the observed value of X. How is this done?

Consider the $n \times p$ matrix $\mathcal{X} = (BX, \ldots, B^p X)$. Equation (11.1.1) becomes

$$X = \mathcal{X}\theta + \epsilon .$$

Thus, X is a point in the space spanned by $\mathcal{X}$ plus a random vector. A reasonable guess for θ is θ_{LS} such that $\mathcal{X}\theta_{LS}$ is the projection of X onto the space spanned by $BX, \ldots, B^p X$. This is called the least square estimator of θ. Whenever $\mathcal{X}$ is of rank p, we have

$$\theta_{LS} = (\mathcal{X}^{\mathrm{T}}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}} X .$$

This can be calculated solely on the observed X.

Furthermore, notice that the (i,j)-entry of the matrix $\mathcal{X}^{\mathrm{T}}\mathcal{X}$ is

$$\begin{aligned}\langle B^i X, B^j X\rangle &= \sum_{1+(i\vee j)\leqslant k\leqslant n} X_{k-i}X_{k-j} \\ &= \sum_{1+|j-i|\leqslant r\leqslant n-(i\wedge j)} X_r X_{r-|i-j|} ,\end{aligned}$$

while the i-th coordinate of $\mathcal{X}^{\mathrm{T}} X$ is $\langle B^i X, X\rangle$.

It is customary to define the empirical autocovariances of order $k < n$ as

$$\gamma_n(k) = n^{-1} \sum_{k+1\leqslant r\leqslant n} X_r X_{r-k} .$$

Notice that

$$\langle B^i X, B^j X\rangle - n\gamma_n(|i-j|) = \sum_{n-(i\wedge j)<r\leqslant n} X_r X_{r-|i-j|} .$$

Therefore, whenever $X_n = O_P(1)$ as n tends to infinity, and $i - j$ is fixed, we have

$$\langle B^i X, B^j X\rangle = n\gamma_n(|i-j|) + O_P(1) \qquad \text{as } n\to\infty . \tag{11.1.2}$$

This explains why the most popular estimator of θ is not θ_{LS} but the following substitute. Define the matrix

$$\Gamma_n = \bigl(\gamma_n(|i-j|)\bigr)_{1\leqslant i,j\leqslant p}$$

and the vector

$$\gamma_n = \big(\gamma_n(i)\big)_{1\leqslant i\leqslant p}\,.$$

If the process X_n is of order 1 as n tends to infinity, then (11.1.2) shows that

$$(n^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{X})^{-1} = \Gamma_n^{-1} + O_P(n^{-1})$$

while

$$\mathcal{X}^{\mathrm{T}}X = \gamma_n + O_P(n^{-1})\,.$$

Thus, instead of using θ_{LS}, one tends to guess θ by

$$\hat\theta_n = \Gamma_n^{-1}\gamma_n\,.$$

Actually, whether we use θ_{LS} or $\hat\theta_n$ does not really matter much for our purposes. What we do care about is that X is a linear function of ϵ as (11.1.1) shows. The autocovariance $\gamma_n(k)$ is a quadratic form in X, as well as in ϵ, a classical fact. It is then plain that tail probabilities of $\gamma_n(k)$ are relevant to statistics, and that chapter 8 provides the right estimates.

Maybe in order to fully enjoy the results we are going to prove, one should know the basics of the classical theory. To make this text quite selfcontained, let us sketch it. To this end, define the polynomial

$$\Theta(z) = 1 - \sum_{1\leqslant i\leqslant p}\theta_i z^i\,,\qquad z\in\mathbb{C}\,.$$

Equation (11.1.1) can be rewritten as

$$\Theta(B)X = \epsilon\,.$$

Denote by $r_1,\ldots,r_p$ the complex roots of Θ. Then $\Theta(z) = \prod_{1\leqslant i\leqslant p}(1-r_i^{-1}z)$. We can write $1/\Theta(z)$ formally as a series. This is done conveniently by introducing the vector $r = (r_1,\ldots,r_p)$. Whenever $s = (s_1,\ldots,s_p)$ belongs to $\mathbb{Z}^p$, we write $|s| = s_1+\cdots+s_p$ and $r^s = \prod_{1\leqslant i\leqslant p} r_i^{s_i}$. We have

$$\begin{aligned}1/\Theta(z) &= \prod_{1\leqslant i\leqslant p}(1-r_i^{-1}z_i)^{-1} = \prod_{1\leqslant i\leqslant p}\sum_{k\geqslant 0} r_i^{-k}z_i^k\\ &= \sum_{k\geqslant 0}\Big(\sum_{|s|=k} r^{-s}\Big)z^k\,.\end{aligned}$$

Substituting B for z, we can formally define $\Theta(B)^{-1}$. Writing $\epsilon_i = 0$ if $i \leqslant s$, one can then check that $X = \Theta(B)^{-1}\epsilon$, that is

$$X_n = \sum_{k\geqslant 0} \Bigl(\sum_{|s|=r} r^{-s} \Bigr) \epsilon_{n-k} \, . \tag{11.1.3}$$

When all the roots r_i are outside the unit circle, there exists a positive η such that

$$\begin{aligned} \Bigl| \sum_{|s|=k} r^{-s} \Bigr| &\leqslant (1+\eta)^{-k} \sharp\{\, s : |s| = k \,\} \\ &= (1+\eta)^{-k} \binom{k+p-1}{p-1} \\ &\sim (1+\eta)^{-k} \frac{k^p}{(p-1)!} \qquad \text{as } k \to \infty \, . \end{aligned}$$

Therefore if the residuals ϵ_i have a tail which decays fast enough the distribution of X_n converges weakly* to that of $\sum_{k\geqslant 0} (\sum_{|s|=k} r^{-s}) \epsilon_k$. A simple condition on the tail of ϵ_i for this convergence to hold is

$$\int_1^\infty P\{\, |\epsilon_i| \geqslant t \,\} \frac{\mathrm{d}t}{t} < \infty \, .$$

A more stringent one is to assume that ϵ_i is integrable; in this case X_n converges in L^1 as well.

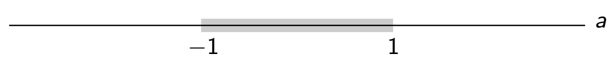
If now some roots are inside the unit disk, we can first assume that r_1 is the unique root with smallest modulus; and therefore $|r_1|$ is less than 1. Then, (11.1.3) shows that the distribution of $r_1^n X_n$ converges weakly* to a nondegenerate limit. Since r_1^n converges to 0 exponentially fast, this amounts to saying that the process X_n explodes at exponential rate. Some complications occur if the smallest root is not unique, but X_n still explodes, essentially at an exponential rate.

As a consequence, the asymptotic behavior of the empirical autocovariances as n tends to infinity is very different according to the location of the roots r_i with respect to the unit disk.

In conclusion, the classical theory makes a great deal of the location of the roots of Θ with respect to the unit disk. And it is essentially all that it cares about, because only the behavior as the time n goes to infinity is considered. In the following sections, we will show that it is only a part of the overall behavior of these processes.

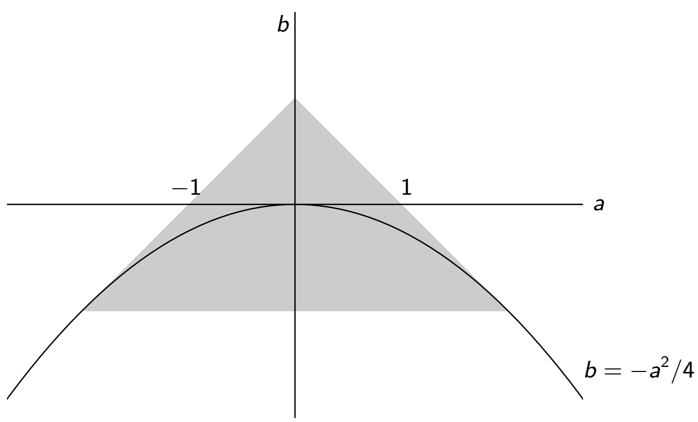
To end this section, let us examine this root question for autoregressive models of order 1 and 2.

For an autoregressive process of order 1, we write $X_i = aX_{i-1} + \epsilon_i$. If $|a| < 1$, this process is nonexplosive. This can be represented on the real line as follows.



Shaded region for the nonexplosive domain

For an autoregressive process of order 2, we write it as $X_i = aX_{i-1} + bX_{i-2} + \epsilon_i$. We need to determine where a, b should lie for the polynomial $x^2 - ax - b$ to have all its roots within the unit disk. If $a^2 + 4b$ is negative, the roots are complex, conjugate to each others. They are in the unit disk if and only if their product is less than 1, that is if $-b < 1$. If $a^2 + 4b$ is nonnegative and a is positive, the largest root in absolute value is $(a + \sqrt{a^2 + 4b})/2$. It is less than 1 if $a^2 + 4b < (2 - a)^2 = a^2 - 4a + 4$, that is $b + a < 1$. One can argue similarly if a is negative, and we obtain the following triangular domain.



Shaded region for the nonexplosive domain

11.2. Autoregressive process of order 1.

In this section, we investigate the tail behavior of the autocovariances of autoregressive processes of order 1,

$$
\begin{aligned}
&X_1 = \epsilon_1 \\
&X_n = aX_{n-1} + \epsilon_n\,, \qquad n \geqslant 2\,.
\end{aligned}
$$

Recall that $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ and $X = (X_1, \ldots, X_n)$. Defining

$$A = \begin{pmatrix} 1 & & & & 0 \\ a & \ddots & & & \\ a^2 & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \\ a^{n-1} & \cdots & a^2 & a & 1 \end{pmatrix},$$

we see that $X = A\epsilon$. The matrix of the backward shift on $\mathbb{R}^n$ is

$$B = \begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}.$$

The empirical covariance of order k,

$$n\gamma_n(k) = \sum_{k+1 \leqslant i \leqslant n} X_i X_{i-k} = \langle X, B^k X \rangle = \langle A^{\mathrm{T}} B^k A\epsilon, \epsilon \rangle$$

is a nice quadratic form in ϵ. If $k \geqslant n$, then $\gamma_n(k) = 0$. We assume from now on that $k < n$. When ϵ has a heavy tail, the tail behavior of $n\gamma_n(k)$ depends on the value of a. It is given by the following result.

11.2.1. THEOREM. *Let X be an autoregressive process of order one, with coefficient a and independent and identically distributed innovations ϵ_i having a Student-like distribution with parameter α. The following expressions are equivalent to $P\{\, n\gamma_n(k) \geqslant t \,\}$ as t tends to infinity.*

(i) If $a = 0$ and $k = 0$,

$$K_{s,\alpha} \alpha^{(\alpha-1)/2} 2n\, t^{-\alpha/2}.$$

(ii) If $a = 0$ and $k \geqslant 1$,

$$K_{s,\alpha} \alpha^{\alpha} 2(n-k)_+\, t^{-\alpha} \log t.$$

(iii) k is even and $a \neq 0$, or k is odd and $a > 0$,

$$K_{s,\alpha} \alpha^{(\alpha-1)/2} 2 a^{k\alpha/2} \sum_{1 \leqslant i \leqslant n-k} \left| \frac{1 - a^{2i}}{1 - a^2} \right|^{\alpha/2} t^{-\alpha/2}.$$

(iv) k is odd and $a < 0$,

$$K_{s,\alpha}^2 \alpha^{\alpha} t^{-\alpha} \log t \times$$

$$\sum_{n-k+1 \leqslant i \leqslant n} \sum_{1 \leqslant j \leqslant n} a^{\alpha|i-k-j|} \left| \frac{2 - a^{2(n+1)} \big(a^{2(i \vee (j+k))} - a^{2(j \vee (i+k))} \big)}{1 - a^2} \right|^{\alpha}$$

It is implicit in the statement that the function $a \mapsto (1-a^i)/(1-a^2)$ is extended by continuity at $a=1$. Its value for $a=1$ is $i/2$.

The striking fact is that odd and even autocovariances exhibit very different decays when a is negative; the former are of order $t^{-\alpha/2}$, the latter of order $t^{-\alpha}\log t$.

Proof. Define the matrix $C = A^{\mathrm{T}}B^kA$. We apply the results of chapter 8. We need to check if the largest diagonal coefficient of C is positive, zero, or negative. In order to calculate it, notice that

$$A_{i,j} = \begin{cases} a^{i-j} & \text{if } i \geqslant j, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(B^k)_{i,j} = \begin{cases} 1 & \text{if } k+1 \leqslant i \leqslant n \text{ and } j = i-k, \\ 0 & \text{otherwise.} \end{cases}$$

Consequently,

$$\begin{aligned} C_{i,i} &= \sum_{1\leqslant j,m\leqslant n} (A^{\mathrm{T}})_{i,m}(B^k)_{m,j}A_{j,i} = \sum_{i+k\leqslant m\leqslant n} A_{m,i}A_{m-k,i} \\ &= \begin{cases} a^k \dfrac{1-a^{2(n-i-k+1)}}{1-a^2} & \text{if } a^2 \neq 1 \text{ and } n-i-k \geqslant 0, \\ a^k(n-i-k+1) & \text{if } a^2 = 1 \text{ and } n-i-k \geqslant 0, \\ 0 & \text{if } i \geqslant n-k+1. \end{cases} \end{aligned}$$

Assume $a=0$ and $k=0$. Then $C = \mathrm{Id}$. Statement (i) of Theorem 11.2.1 follows from Theorem 8.2.1.

If a is null and k is nonzero, the matrix $C = B^k$ has all its diagonal elements vanishing. We apply Theorem 8.3.1, calculating

$$\sum_{i:C_{i,i}=0}\sum_{1\leqslant j\leqslant n} |(B^k)_{i,j} + (B^k)_{j,i}|^\alpha = 2(n-k)_+ \,.$$

When k is even and a is nonzero, then $C_{i,i}$ is positive for any $1 \leqslant i \leqslant n-k$. We apply Theorem 8.2.1, calculating

$$\sum_{1\leqslant i\leqslant n-k} C_{i,i}^{\alpha/2} = a^{k\alpha/2} \sum_{1\leqslant i\leqslant n-k} \Big(\frac{1-a^{2(n-i-k+1)}}{1-a^2}\Big)^{\alpha/2} .$$

This gives statement (iii), after substituting $n-i-k+1$ for i in the summation.

Statement (iii), when k is odd and a is positive, follows from exactly the same calculation.

Let us now concentrate on k odd and a negative. Then a^k is negative, and so is $C_{i,i}$ if $1 \leqslant i \leqslant n-k$, while $C_{i,i}$ vanishes if $i \geqslant n-k+1$. Therefore, we apply Theorem 8.3.1. We need to calculate

$$\sum_{n-k+1\leqslant i\leqslant n} \sum_{1\leqslant j\leqslant n} |C_{i,j}+C_{j,i}|^{\alpha} .$$

We have

$$\begin{aligned} C_{i,j} &= \sum_{1\leqslant l,m\leqslant n} (A^{\mathrm{T}})_{i,l}(B^k)_{l,m}A_{m,j} = \sum_{i\vee(j+k)\leqslant l\leqslant n} a^{l-i}a^{l-k-j} \\ &= a^{|i-j-k|}\frac{1-a^{2(n+1-(i\vee(j+k)))}}{1-a^2} . \end{aligned}$$

We obtain $C_{j,i}$ by permuting i and j. This gives statement (iv). ∎

How good are these approximations? Looking at the bound in Theorem 3.1.9 and how we derived Theorem 5.1, we cannot expect them to be good when we are integrating in a high dimensional space, that is when n is large.

A plot of the approximations given in Theorem 11.2.1 does not show much, since all the probabilities go to 0 as t tends to infinity. When comparing the tails, it makes more sense to look at the relative error. This leads to the plot

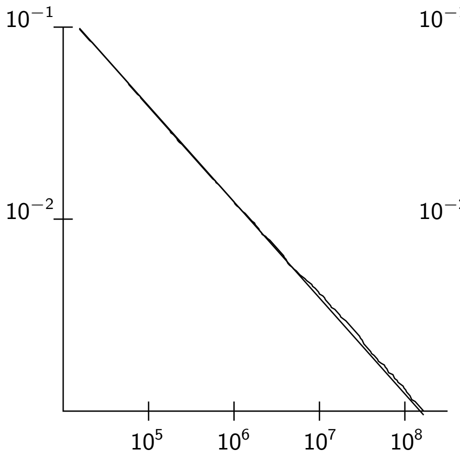
$$\log P\{\, n\gamma_n(k)\geqslant t\,\}$$

as well as the logarithm of the approximation. These should be approximately in linear relation with $\log t$. Therefore, the plots below will show the function $t\mapsto P\{\,\gamma_n(k)\geqslant t\,\}$ with both axes in logarithmic scale. Since we do not know a closed formula for $P\{\,n\gamma_n(k)\geqslant t\,\}$, we obtained this probability by simulation. We generated 100,000 replicas of ϵ. As t increases, the estimate of the true probability is based on less and less points; the simulated curve tends to wiggle as t gets large. The theoretical approximation will be the smooth curve on the graphs. The parameters involved are k, n, α, a. We will only consider the autocovariance of order 1 in our simulations. We consider the sample sizes $n=10$, which is very small, and $n=20$, which is a common order of magnitude in some applications. We also consider probabilities of interest in applications, namely between 10^{-1} and 10^{-3}. Recall that 5% is about $10^{-1.30}$.

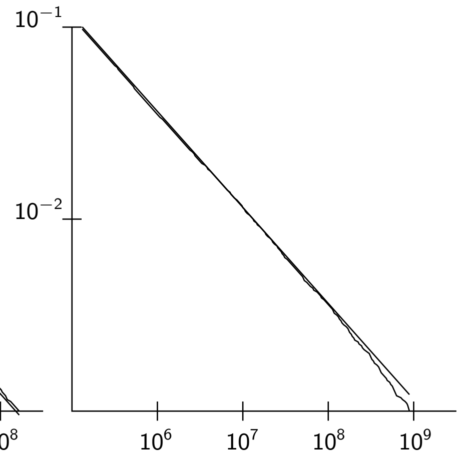
The results are as follows.

Let us first see what happens when the errors have a Cauchy distribution, corresponding to $\alpha = 1$.

When $a = 1$, the two plots bellow show that the approximation is amazingly good.

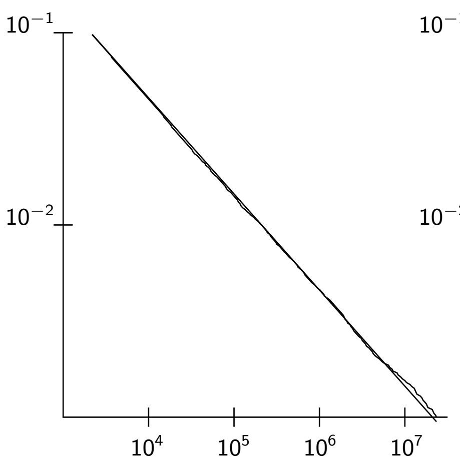


$n = 10$, $a = 1$, errors Cauchy

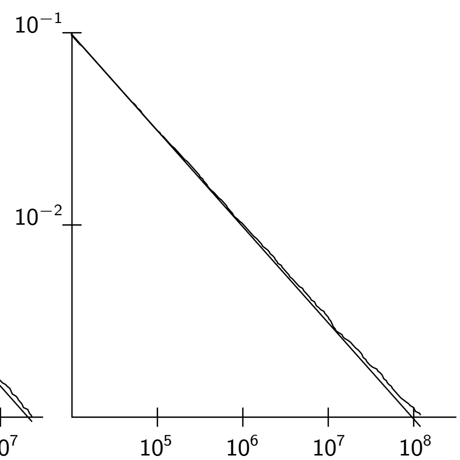


$n = 20$, $a = 1$, errors Cauchy

For $a = 0.5$ the approximation is also excellent.

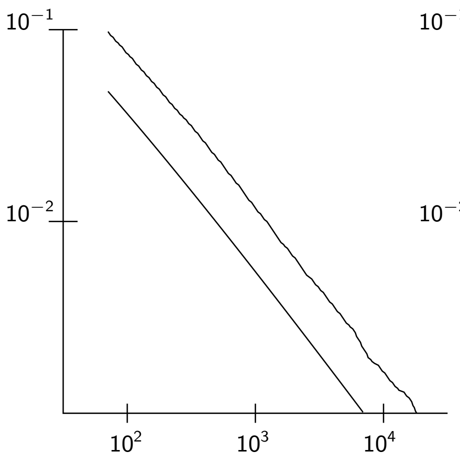


$n = 10$, $a = 0.5$, errors Cauchy

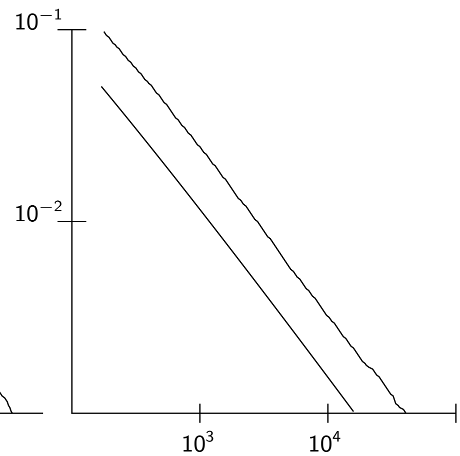


$n = 20$, $a = 0.5$, errors Cauchy

In the degenerate case where we need to apply Theorem 8.3.1, the coefficient a vanishes. The approximation is not as good as before. But taking into account that we are visualizing a relative error, it performs well enough to be of practical use.

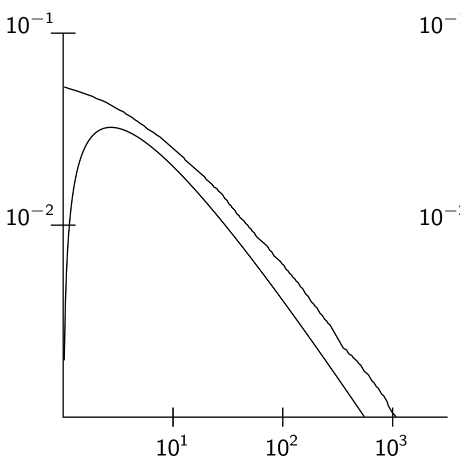


$n = 10,\ a = 0$, errors Cauchy

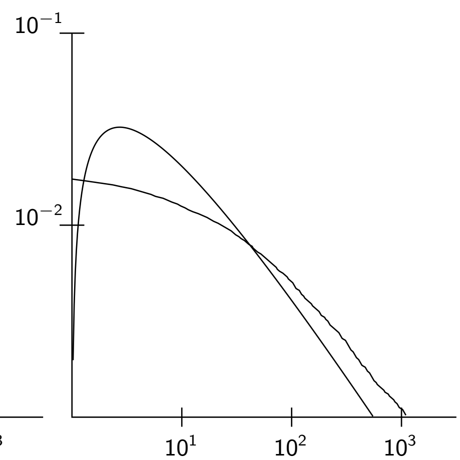


$n = 20,\ a = 0$, errors Cauchy

For $a = -0.5$, the autocovariance tends to be negative since $\gamma_n(1)/\gamma_n(0)$ is an approximation of a. Therefore we need to go further on the tail to have a good approximation. For $n = 10$, it is still accurate enough to be of some practical interest. One can use the approximation to find critical values at levels less than $10^{-1.5} \approx 3\%$ say. But as n increases from 10 to 20, the accuracy decreases.

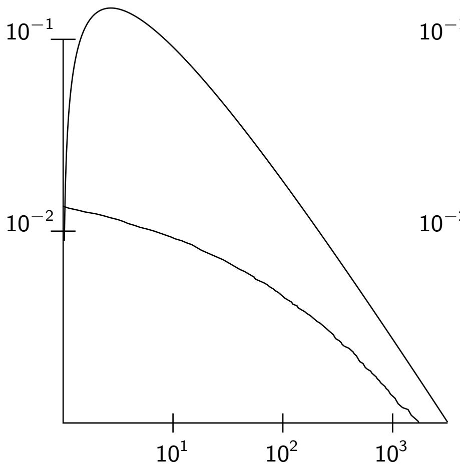


$n = 10,\ a = -0.5$, errors Cauchy

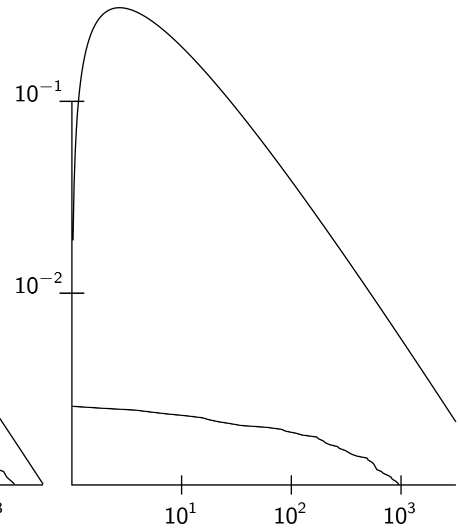


$n = 20,\ a = -0.5$, errors Cauchy

For $a = -1$, we need to go much further in the tail of the distribution of $\gamma_n(1)$ in order to have a positive quantile. The approximation is not accurate in the range of practical interest. As n increases from 10 to 20, the approximation cannot be used in applications.



$n = 10$, $a = -1$, errors Cauchy



$n = 20$, $a = -1$, errors Cauchy

When a is negative, it makes more sense to approximate the lower tail. Given what we have done, it is a trivial matter. We state a result only in the form needed for our discussion.

11.2.2. THEOREM. *Let X be an autoregressive process of order 1, with coefficient a and errors independent and identically distributed from a Student-like distribution with parameter α. If a is negative, then*

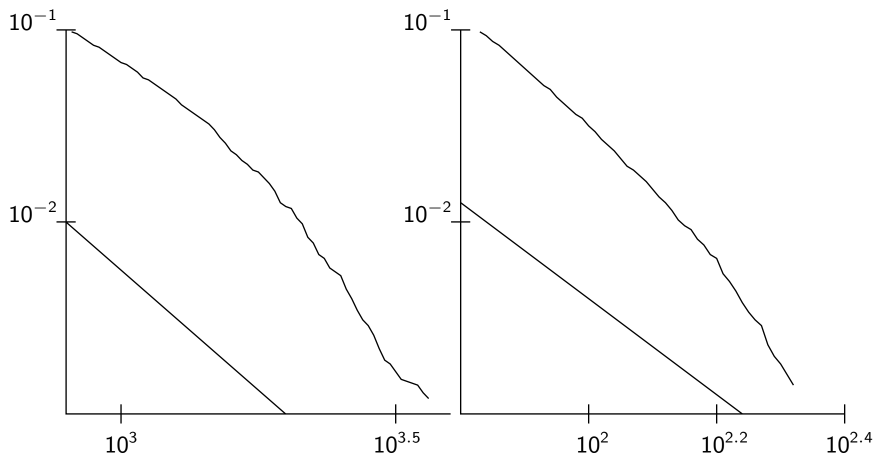
$$P\{\, n\gamma_n(1) \leqslant -t \,\} \sim K_{s,\alpha}\alpha^{(\alpha-1)/2} 2a^{\alpha/2} \sum_{1\leqslant i\leqslant n-1} \Big(\frac{1-a^{2j}}{1-a^2}\Big)^{\alpha/2} t^{-\alpha/2}$$

as t tends to infinity.

Proof. With the notation of the proof of Theorem 11.2.1, we need to evaluate the upper tail of $\langle -C\epsilon, \epsilon\rangle$. In the proof of Theorem 11.2.1, we shown that when $k = 1$ and a is negative, the matrix $-C$ has its coefficients $C_{n,n}$ vanishing, while $C_{i,i} = a(1-a^{2(n-i)})/(1-a^2)$ for $i = 1, \ldots n-1$. Apply Theorem 8.2.1 to conclude the proof. ∎

There is not much point in reproducing here results on the approximation of the lower tail. It is enough to say that it works as expected; that is, the approximation is very sharp when $a = -1$ or $a = -0.5$. The pictures look identical to those for the upper tail with positive coefficient a.

As α increases, the approximation given in Theorem 11.2.1 degenerates for positive values of a. For a Student-distribution with 5 degrees of freedom, that is $\alpha = 5$, and $n = 10$, their use starts to be questionable.

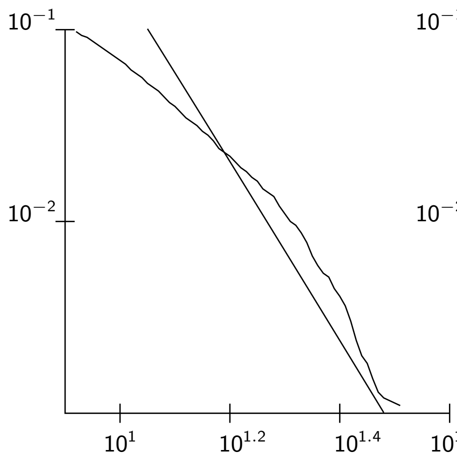


$n = 10$, $a = 1$, errors Student(5) $n = 20$, $a = 1$, errors Student(5)

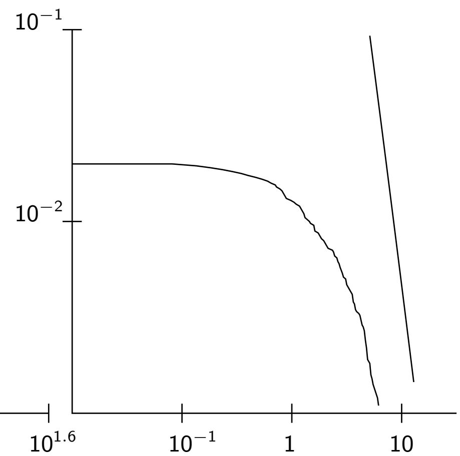
For $a = 0$, we appealed to Theorem 8.3.1, and the approximation is not so good. We would like to point out that this failure can be seen on a very simple example. Consider two independent random variables, X, Y, with density $\alpha/x^{\alpha+1}$ over $[1, \infty)$. We can calculate explicitly the distribution of their product,

$$\begin{aligned} P\{ XY \geqslant t \} &= \alpha \int_{x \geqslant 1} x^{-\alpha-1} P\{ Y \geqslant t/x \} dx \\ &= \alpha \int_{x \geqslant 1} x^{-\alpha-1} \big((x/t)^{\alpha} \wedge 1 \big) dx \\ &= \frac{\alpha \log t}{t^{\alpha}} + \frac{1}{t^{\alpha}} \, . \end{aligned}$$

Our approximation picks up the leading term, $\alpha t^{-\alpha} \log t$ as t tends to infinity. But the growth of $\log t$ is too slow for the first term to really dominate in the range where the probability is of order 10^{-1} or 10^{-2}. One cannot expect a good one-term approximation in such case, except if α is large. This suggests that when a is zero, our approximation may improve when α increases. This is the case. And of course, for negative a, it becomes worse, positive values of the empirical covariance being less and less likely.

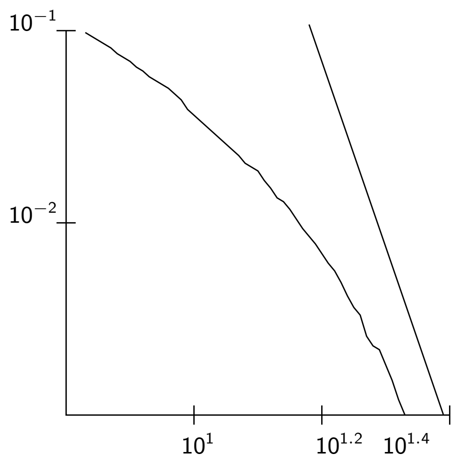


$n = 20$, $a = 0$, errors Student(5)



$n = 20$, $a = -0.5$, errors Student(5)

For $a = 0$, the approximation is fairly good, until we have many moments on the distribution, that is if α is large enough. Then assuming simply that the errors are normally distributed may give a better approximation.



$n = 20$, $a = 0$, errors Student(10)

When our approximation is not so good, one can think of some alternative techniques. Besides the classical Edgeworth expansion — which is poor in term of relative error — one could also approximate the Student distribution by a normal one, and then proceed as if ϵ were normally distributed. This works well if the normal distribution has the "right" variance. But one has to be aware that for α small, the right variance *is not* that of the corresponding Student distribution.

For instance, for the Student distribution with 5 degrees of freedom, the normal approximation using a variance equal to that of the Student distribution poor. One needs a much larger variance. I tried to approximate the Student distribution by a normal with the variance such that some quantile of the normal would be equal to that of the Student. This does not work any better, in the sense that there is no systematic way to do this kind of calibration. One should also be aware that the symmetry in the Student distribution makes our approximations less precise. They would be more accurate if the errors had a centered Pareto distribution for instance. Obviously more work is needed to derive a set of approximations which would cover more or less any regime. It is doubtful that a single approximation scheme can give satisfactory results under a very broad class of distributions for the errors and relatively arbitrary sample size.

Classically, a is estimated by $\hat{a} = \gamma_n(1)/\gamma_n(0)$. Let us now consider the test problem

$$\mathrm{H}_0 \,:\, a \leqslant a_0, \qquad \text{versus} \qquad \mathrm{H}_1 \,:\, a > a_0.$$

A possible way to perform this test is to reject the null hypothesis if $\hat{a}_n - a_0$ is too large, that is if $\gamma_n(1) - a_0\gamma_n(0)$ is too large. For a reason which will be explained in the proof of the next result, it is better to use

$$\hat{\gamma}_n(0) = n^{-1} \sum_{1 \leqslant i \leqslant n-1} X_i^2$$

instead of $\gamma_n(0)$. As n tends to infinity, the classical theory ensures that it does not make any difference. However, it does make a difference for our finite sample results. The theory is much nicer with $\hat{\gamma}_n(0)$.

The test statistics is again a quadratic form in ϵ. The following result gives its tail approximation under the null as well as under the alternative hypothesis. As one more parameter is involved, another behavior appears.

11.2.2. THEOREM. *Let X be an autoregressive process of order* 1*, with coefficient a and independent and identically distributed innovations, all having a Student-like distribution with parameter α. The tail probability*

$$P\big\{\, n\big(\gamma_n(1) - a_0\hat{\gamma}_n(0)\big) \geqslant t \,\big\}$$

admits the following equivalent as t tends to infinity.

(i) If $a > a_0$,

$$K_{s,\alpha}\alpha^{\frac{\alpha-1}{2}}2(a-a_0)^{\alpha/2}\sum_{1\leqslant i\leqslant n-1}\Bigl(\frac{1-a^{2(n-i)}}{1-a^2}\Bigr)^{\alpha/2}t^{-\alpha/2}\,.$$

(ii) If $a_0 = a$ and both are nonnegative,

$$K_{s,\alpha}^2\alpha^\alpha 2\sum_{1\leqslant k\leqslant n-2}(n-k)a^{(k-1)\alpha}\,t^{-\alpha}\log t\,,$$

with the convention that $0^0 = 1$ when $a = 0$.
(iii) If a_0 is positive and $a < a_0$,

$$c(a_0,a,\alpha,n)\,t^{-\alpha}$$

for some function $c(\cdot)$.

In case (iii), we will explain after the proof how to calculate the function $c(\cdot)$ in a typical case.

Proof. Write $n\bigl(\gamma_n(1)-a_0\hat\gamma_n(0)\bigr) = \epsilon^{\mathrm{T}}C\epsilon$ with

$$C = A^{\mathrm{T}}BA - a_0A^{\mathrm{T}}B^{\mathrm{T}}BA\,.$$

From the proof of Theorem 11.2.1 we obtain the diagonal terms $(A^{\mathrm{T}}BA)_{i,i}$. We calculate

$$(A^{\mathrm{T}}B^{\mathrm{T}}BA)_{i,i} = \sum_{1\leqslant k\leqslant n}(BA)_{k,i}^2 = \sum_{1\leqslant k\leqslant n-1}A_{k,i}^2 = \frac{1-a^{2(n-i)}}{1-a^2}\,.$$

Consequently,

$$C_{i,i} = \begin{cases}(a-a_0)\dfrac{1-a^{2(n-i)}}{1-a^2} & \text{if } 1\leqslant i\leqslant n-1\\ 0 & \text{if } i=n.\end{cases}$$

If instead of using $\hat\gamma_n(0)$ we use $\gamma_n(0)$, the term $(A^{\mathrm{T}}B^{\mathrm{T}}BA)_{i,i}$ has denominator $1-a^{2(n-i+1)}$. The discussion is a bit more involved. The result becomes more dependent on n. It is more complicated to state, but it does not make much difference as far as the theory goes.

If $a-a_0$ is positive, then all the diagonal coefficients of C but $C_{n,n}$ are positive. We apply Theorem 8.3.1.

If $a - a_0$ is negative, only $C_{n,n}$ is nonnegative. We apply Theorem 8.3.1. Since we will need it, let us calculate $C_{i,j}$. First

$$(BA)_{k,i} = \begin{cases} A_{k-1,i} & \text{if } 2 \leqslant k \leqslant n, \\ 0 & \text{if } k = 1. \end{cases}$$

Therefore,

$$(BA)_{k,i} = \begin{cases} a^{k-1-i} & \text{if } 2 \leqslant k \leqslant n, \\ 0 & \text{otherwise.} \end{cases}$$

Consequently, a little algebra shows that

$$(A^{\mathrm{T}} B^{\mathrm{T}} BA)_{i,j} = \sum_{1 \leqslant k \leqslant n} (BA)_{k,i} (BA)_{k,j} = a^{|i-j|} \frac{a^{2n} - a^{2(i \vee j)}}{a^2 - 1} .$$

In particular, for $i < n$,

$$C_{n,i} = a^{n-i-1},$$

while $C_{i,n} = 0$ for $1 \leqslant i \leqslant n$. Therefore

$$\sum_{1 \leqslant i \leqslant n} |C_{i,n} + C_{n,i}|^\alpha = \sum_{1 \leqslant i \leqslant n} a^{\alpha(n-i-1)} = \sum_{0 \leqslant i \leqslant n-2} a^{\alpha^i} .$$

If a and a_0 are equal, then all the diagonal coefficients of C vanish. In this case, for i, j distinct,

$$C_{i,j} = a^{|i-j-1|} \frac{1 - a^{2(n+1-(i \vee (j+1)))}}{1 - a^2} - a_0 a^{|i-j|} \frac{a^{2n} - a^{2(i \vee j)}}{a^2 - 1} .$$

If a_0 is positive and strictly larger than a, then all the diagonal coefficients of C are negative. We need to determine $N(C)$ and apply Theorem 8.2.10. The statement follows from Theorem 8.2.21. ■

Some useful information can be deduced from Theorem 11.2.2. A first qualitative deduction is that if a_0 is positive under the null hypothesis, then the tail probability under consideration has a very different decay according to the position of a with respect to a_0. Thus, one should probably not use symmetric confidence intervals. It may be wise to have a somewhere on the right half of the interval.

Next, assume that we want to test with the risk of first type η very small. We can use the approximation under the null hypothesis to obtain the critical value. Define

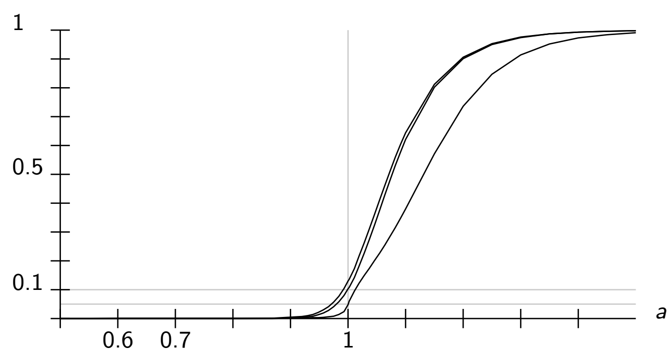
$$c(a) = 2K_{s,\alpha} \alpha^{\frac{\alpha-1}{2}} (a - a_0)^{\alpha/2} \sum_{1 \leqslant i \leqslant n-1} \Big(\frac{1 - a^{2(n-i)}}{1 - a^2} \Big)^{\alpha/2} .$$

We see that if $0 \leqslant a_0 < a$, then $\eta \sim c(a)/t^{\alpha/2}$. This gives an approximate critical value $t_\eta = \big(c(a)/\eta\big)^{2/\alpha}$.

The following plot shows the actual risk of the first type when using the approximate critical value. The "true" value of the risk is obtained by simulation, replicating 100,000 copies of the vector ϵ. The curves were obtained by linearly interpolating between the following values for a.

0.500	0.552	0.605	0.657	0.710	0.762	0.815	0.868	0.920	0.930
0.939	0.948	0.957	0.966	0.975	0.984	0.993	1.000	1.002	1.011
1.019	1.028	1.037	1.046	1.055	1.064	1.073	1.082	1.091	1.100
1.150	1.200	1.250	1.300	1.350	1.400	1.450	1.500		

The sample size is $n = 20$. The lowest curve is for $\alpha = 1$, the two almost equal curves are for $\alpha = 5$ — the lower one of the two curves — and $\alpha = 10$. The value $a = 1$ is indicated, as well as the levels 5% and 10% .



$n = 20,\ \alpha = 1,\ 5,\ 10$

The result is satisfying at first glance. A more careful examination shows some reason to worry. For $a = 1$, the actual risk of the test is about 10% when $\alpha = 5$ or 10. This is still small, but in terms of relative error, this is twice as much as what we wanted. The fact that the power function grows moderately fast with a is not a surprise if you plot some of these processes. On a trajectory of length 20, and with errors having a Cauchy distribution, an autoregressive process of order one with $a = 1$ looks very similar to one with $a = 1.2$ for instance.

11.3. Autoregressive processes of arbitrary order.

In principle, all the results of the previous section can be generalized to autoregressive processes of arbitrary order. As more parameters are

involved in the model, the analysis is harder. Thus, our goal is rather modest. We will prove that the tail of the autocovariance of order 1 has different decays according to the values of the parameters, and the number of observations as well. Since autoregressive processes of order one are a degenerate case of those of higher order, the higher order autocovariances would have an even more complex behavior. For any given value of the parameters, the results of chapter 8 can be used to do numerical computations; this is quite easy, and sometimes helpful, but does not provide further insights. For autoregressive processes of order 2, we will obtain some rather precise results, showing the intricacy of these models.

Let us now consider an autoregressive model of order p as in (11.1.1). The parameter $\theta = (\theta_1, \ldots, \theta_p)$ is in $\mathbb{R}^p$. Our first result shows that $\mathbb{R}^p$ can be partitioned into two regions, one where the tail behavior of $n\gamma_n(1)$ is like $t^{-\alpha/2}$, the other one where it is like $t^{-\alpha}\log t$. The noticeable fact is that there cannot be other tail behavior, and these regions are nested when the number of observations varies.

11.3.1. THEOREM. *Consider an autoregressive process of order p with errors having a Student-like distribution with parameter α. There exist nonempty semialgebraic sets R_k, $k \geqslant 1$, of $\mathbb{R}^p$, and a positive function $c(\cdot)$ on $\mathbb{R}^p$, such that*

$$P\{\, n\gamma_n(1) \geqslant t \,\} \sim \begin{cases} c(\theta)t^{-\alpha/2} & \text{if } \theta \in \bigcup_{1\leqslant k\leqslant n-1} R_k, \\ c(\theta)t^{-\alpha}\log t & \text{if } \theta \notin \bigcup_{1\leqslant k\leqslant n-1} R_k. \end{cases}$$

REMARK. The regions R_i depend of course on the dimension p of the parameter space. It is remarkable that they do not depend on n, though the quadratic form representing $n\gamma_n(1)$ does. As n increases, $\bigcup_{1\leqslant k\leqslant n} R_k$ increases, but we will see during the proof that its limit is a proper nonempty subset of $\mathbb{R}^p$. The proof also gives some indications on how to calculate these regions efficiently.

Proof of Theorem 11.3.1. We write $X = A\epsilon$. The matrix A is more involved than in section 2. Let e_i be the i-th vector from the canonical basis of $\mathbb{R}^d$, that is having all its entries vanishing, except the i-th one being 1. We denote $A_{i,\bullet}$ the i-th row of A. Relation (11.1.1) gives

$$\begin{aligned} A_{1,\bullet} &= e_1 \\ A_{2,\bullet} &= \theta_1 A_{1,\bullet} + e_2 \end{aligned}$$

$$\vdots$$

$$A_{p+1,\bullet} = \theta_1 A_{p,\bullet} + \theta_2 A_{p-1,\bullet} + \cdots + \theta_p A_{1,\bullet} + e_p \qquad (11.3.1)$$

and for $p+1 \leqslant k \leqslant n$,

$$A_{k,\bullet} = \theta_1 A_{k-1,\bullet} + \theta_2 A_{k-2,\bullet} + \cdots + \theta_p A_{k-p,\bullet} + e_k \,. \qquad (11.3.2)$$

Let $C = A^{\mathrm{T}} B A$ be the matrix of the quadratic form such that $n\gamma_n(1) = \epsilon^{\mathrm{T}} C \epsilon$. Since $A_{j,i} = 0$ for any $1 \leqslant j < i \leqslant n$,

$$C_{n,n} = 0 \qquad \text{and} \qquad \max_{1\leqslant i\leqslant n} C_{i,i} \geqslant 0 \,.$$

Thus, Theorems 8.2.1 and 8.3.1 show that the only possible tail behaviors of $n\gamma_n(1)$ are either like $t^{-\alpha/2}$ or like $t^{-\alpha}\log t$. They show as well the existence of the nonvanishing function $c(\cdot)$.

Since $A_{i,j} = A_{i+1,j+1}$ for all $i, j \leqslant n-1$, we have for $k \geqslant 1$,

$$\begin{aligned} C_{n-k,n-k} &= \sum_{1\leqslant i,j\leqslant n} (A^{\mathrm{T}})_{n-k,i} B_{i,j} A_{j,n-k} \\ &= \sum_{2\leqslant j\leqslant n} A_{j,n-k} A_{j-1,n-k} \\ &= \sum_{n-k+1\leqslant j\leqslant n} A_{j,n-k} A_{j-1,n-k} \\ &= \sum_{1\leqslant j\leqslant k} A_{n-k+j,n-k} A_{n-k+j-1,n-k} \\ &= \sum_{1\leqslant j\leqslant k} A_{j+1,1} A_{j,1} \,. \end{aligned} \qquad (11.3.3)$$

Relations (11.3.1) and (11.3.2) show that $A_{j,1}$ is a polynomial in a and b. Hence $C_{n-k,n-k}$ is also a polynomial in a and b. The region

$$R_k = \{\, (a,b) \in \mathbb{R}^2 \,:\, C_{n-k+1,n-k+1} \geqslant 0 \,\}$$

is then a semialgebraic set, which does not depend on n. Moreover, the largest diagonal coefficient of C is positive if and only if (a,b) is in $\bigcup_{1\leqslant k\leqslant n} R_k$. The result then follows from Theorems 8.2.1 and 8.3.1. ∎

In practice, to decide in which region we are, we can numerically compute the diagonal terms of C with formulas (11.3.1)–(11.3.3). But if one ultimately wants the constant $c(a,b)$, other elements of C are needed when the tail is like $t^{-\alpha}\log t$.

The sets R_k do not seem easy to describe in general. For applications, this may not be so important, since numerical computation is easy to implement. Understanding their geometry amounts to understanding the behavior of the roots of inductively defined polynomials in the p variables $\theta_1, \ldots, \theta_p$. More can be said when $p = 2$, because polynomials of degree 2 are well understood. And this is enough to show how intricate these autoregressive models are. Thus, from now on, we focus on autoregressive models of order 2. We change slightly the notation, using (a, b) instead of (θ_1, θ_2). Thus, our model is

$$\begin{aligned} X_1 &= \epsilon_1 \,, \\ X_2 &= aX_1 + \epsilon_2 \,, \\ X_k &= aX_{k-1} + bX_{k-2} + \epsilon_k \,, \qquad 3 \leqslant k \leqslant n \,. \end{aligned}$$

We can explicitly write down R_1, R_2, R_3, R_4. Indeed, $C_{n,n} = 0$ and thus $R_1 = \emptyset$. We then have

$$C_{n-1,n-1} = A_{2,1}A_{1,1} = a \,,$$

thus

$$R_2 = \{\, (a, b) \,:\, a > 0 \,\} \,.$$

Since $A_{3,1}A_{2,1} = (a^2 + b)a$,

$$C_{n-2,n-2} = (a^2 + b)a + a = a(a^2 + b + 1) \,.$$

Consequently,

$$R_3 = \{\, (a, b) \,:\, a > 0 \text{ and } b > -1 - a^2 \,;\ \text{ or } a < 0 \text{ and } b < -1 - a^2 \,\} \,.$$

But what matters more,

$$R_2 \cup R_3 = \big((0, \infty) \times \mathbb{R}\big) \cup \{\, (a, b) \,:\, a < 0 \,,\, b < -1 - a^2 \,\} \,.$$

The next figure shows these regions.

We also obtain $A_{4,1} = a(a^2 + 2b)$, which leads to

$$\begin{aligned} C_{n-3,n-3} &= a(a^2 + 2b)(a^2 + b) + a(a^2 + b + 1) \\ &= a\big(2b^2 + b(3a^2 + 1) + a^4 + a^2 + 1\big) \end{aligned}$$

To obtain $R_2 \cup R_3 \cup R_4$, we need to see when $C_{n-3,n-3}$ is positive as a is negative. In other words, when a is negative and

$$2b^2 + b(3a^2 + 1) + 1 + a^2 + a^4 < 0 \,.$$

This holds if $a < \dfrac{1-2\sqrt{2}}{2}$ and

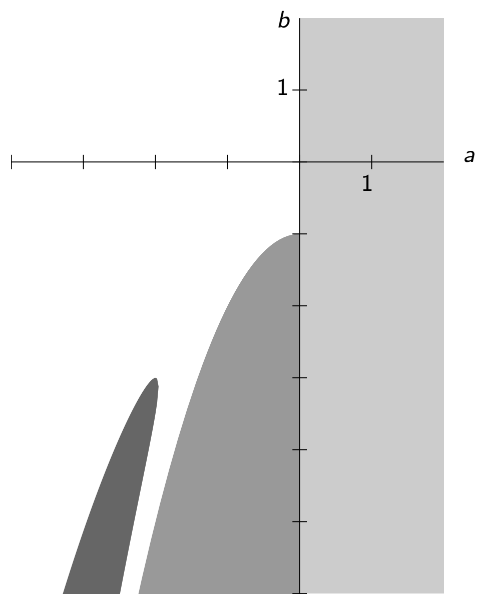
$$b_- \leqslant b \leqslant b \leqslant b_+$$

with

$$b_\pm = \frac{-(1+3a^2) \pm \sqrt{a^4 - 2a^2 - 7}}{4} \,.$$

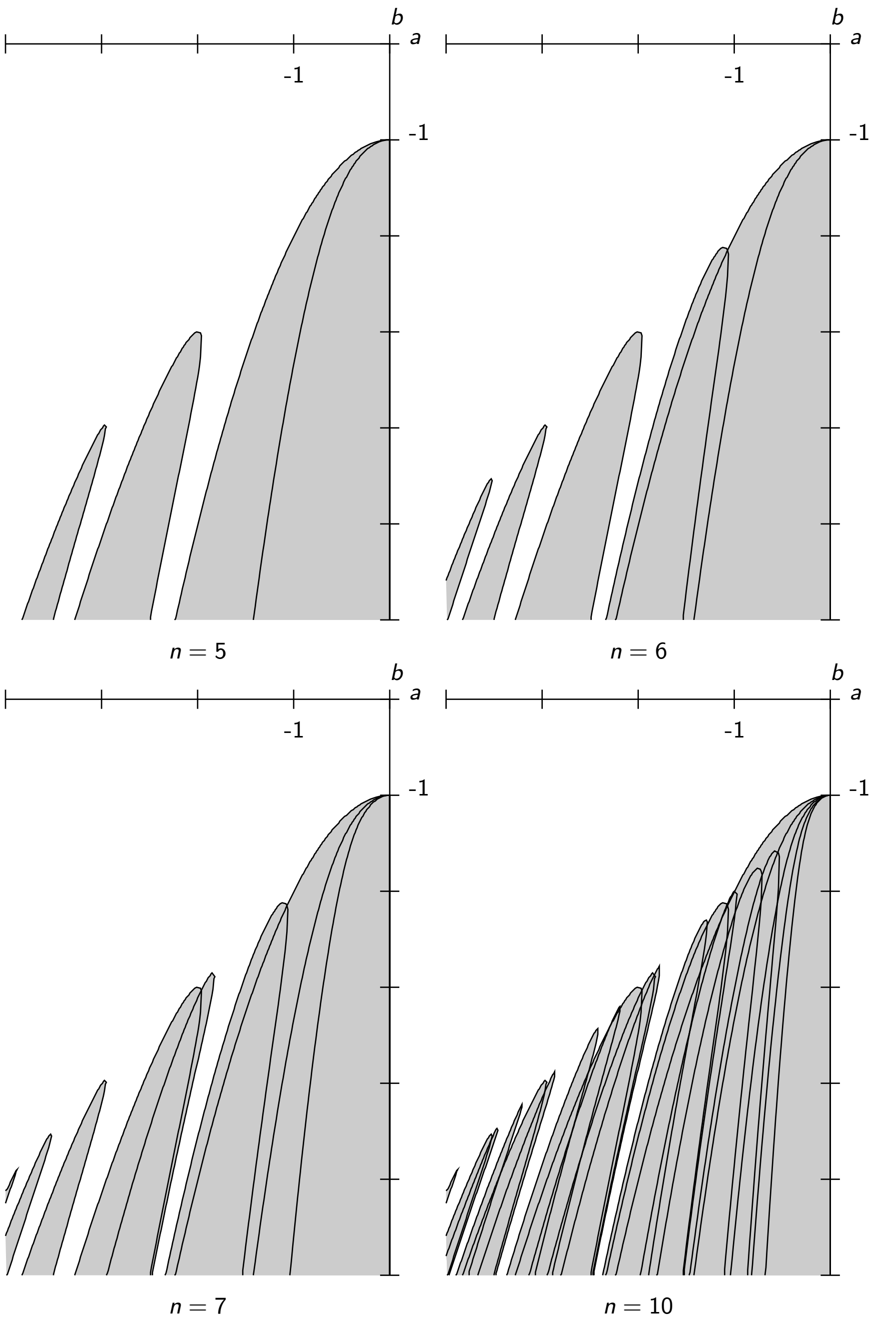
Thus,

$$R_1 \cup R_2 \cup R_3 \cup R_4 = \big((0,\infty)\times\mathbb{R}\big) \bigcup \{\,(a,b)\ :\ a<0\,,\ b<-1-a^2\,\} \bigcup \{\,(a,b)\ :\ a \leqslant \frac{1-2\sqrt{2}}{2}\,;\ b_- \leqslant b \leqslant b_+\,\}\,.$$



Regions R_2, $R_2 \cup R_3$, $R_2 \cup R_3 \cup R_4$.

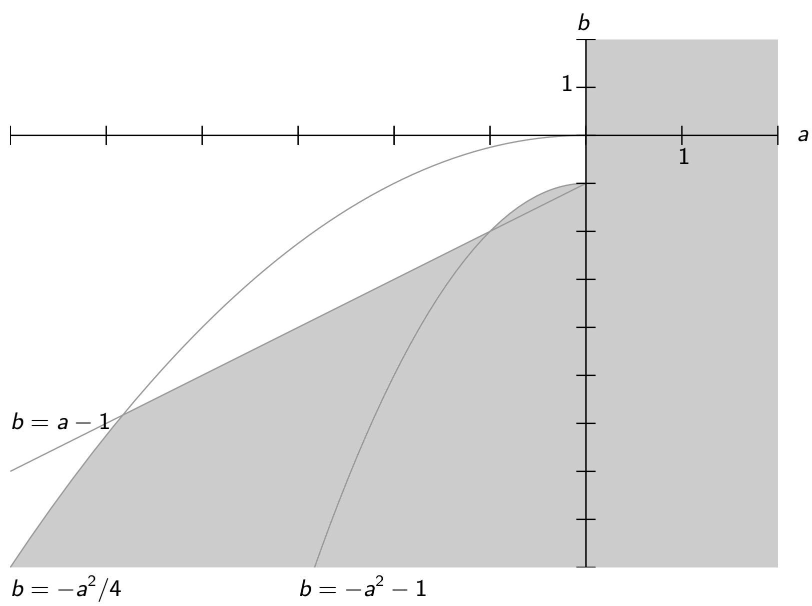
To calculate R_5, we need to solve a cubic equation in b. Closed form expressions are getting more and more cumbersome, and even nonexistent. The following picture shows $\bigcup_{1\leqslant k\leqslant n} R_k$ for $n = 5, 6, 7$ and $n = 10$ in the domain a negative.



These pictures can be more or less understood theoretically. This is the purpose of the next result. Its proof contains even more information, some of it being important, and we will discuss further after the proof.

11.3.2. THEOREM. *The closure of* $\bigcup_{k\geqslant 1} R_k$ *contains all points* (a,b) *for which one of the following conditions holds:*
(i) $a>0$,
(ii) $a\leqslant 0$ *and* $b<-a^2-1$,
(iii) $a\leqslant 0$ *and* $b<\min(-a^2/4, -a-1)$.

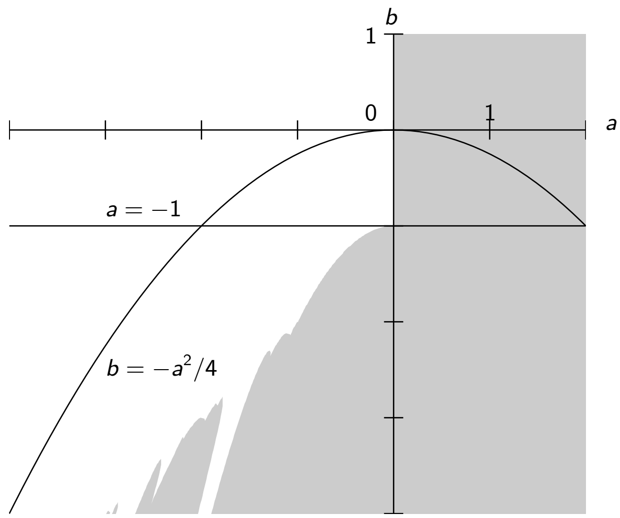
The region described by the three condition in Theorem 11.3.2 is shaded gray in the following picture. It does not contain its boundary.



It follows from Theorems 11.3.2 and 11.3.1 that if (a,b) lies in the gray shaded region, the tail behavior of $P\{\, n\gamma_n(1)\geqslant t\,\}$ is typically like $t^{-\alpha/2}$ for n large enough. I believe that in the nonshaded domain, the tail behavior is like $t^{-\alpha}\log t$; we will prove this only when $b\geqslant a^2/4$.

Proof of Theorem 11.3.2. The proof will be done in examining different regions. We will need several lemmas, and will actually prove much more than the statement.

11.3.3. LEMMA. *If a is positive, so is the largest diagonal coefficient of C. If a is nonpositive, then the largest diagonal coefficient of C vanishes. Consequently,* $\bigcup_{k\geqslant 1} R_k$ *contains the region* $a>0$, *but does not intersect the region* $a\leqslant 0$ *and* $b\geqslant -a^2/4$.



Proof. If a is positive, then $C_{n-1,n-1} = a$ is positive too. To see what is happens when a is nonpositive, denote by u, v the roots of the characteristic equation $x^2 - ax - b = 0$. Equations (11.3.1)–(11.3.2) yield

$$A_{i,j} = ru^{i-j} + sv^{i-j}\,, \qquad i \geqslant j\,,$$

with initial condition $A_{i,i} = 1$ and $A_{i+1,1} = a$. Thus r and s are determined by

$$r + s = 1 \qquad \text{and} \qquad ru + sv = a\,.$$

If u and v are distinct, that is $b \neq -a^2/4$,

$$A_{j,1} = \frac{u^j - v^j}{u - v}\,.$$

If $b = -a^2/4$, we find

$$A_{j,1} = j(a/2)^{j-1}\,.$$

Consequently, if $b \neq -a^2/4$,

$$A_{j+1,1}A_{j,1} = \frac{(u^{j+1} - v^{j+1})(u^j - v^j)}{(u - v)^2}\,. \qquad (11.3.4)$$

Notice that

$$(u - v)^2 = (u + v)^2 - 4uv = a^2 + 4b\,.$$

Let us now assume that a is negative and b is positive. Then a^2+4b is positive. There is no loss of generality in assuming $u<0<v$ since the product of the roots, $-b$, is negative. The inequality

$$u^2-v^2=(u+v)(u-v)=-a\sqrt{a^2+4b}\geqslant 0$$

forces $|u|\geqslant|v|$. Consequently, the sign of $A_{j+1,1}A_{j,1}$ is that of

$$\big((-1)^{j+1}|u|^{j+1}-v^{j+1}\big)\big((-1)^j|u|^j-v^j\big)\,,$$

which is negative. Therefore, the sequence $k\mapsto C_{n-k,n-k}$ is decreasing and

$$\max_{1\leqslant k\leqslant n} C_{k,k}=C_{n,n}=0\,.$$

If we now assume that a is negative and $-a^2/4<b\leqslant 0$, the roots u,v are still real, but both are negative. Thus

$$A_{j+1,1}A_{j,1}=-\frac{(|u|^{j+1}-|v|^{j+1})(|u|^j-|v|^j)}{a^2+4b}$$

is nonpositive since the function $x\mapsto x^j$ and $x\mapsto x^{j+1}$ are increasing on $\mathbb{R}^+$. The sequence $k\mapsto C_{n-k,n-k}$ is decreasing and its maximum is $C_{n,n}=0$. This gives Lemma 11.3.3. ∎

The region left is $a\leqslant 0$ and $b<-a^2/4$. Our next lemma covers a part of it.

11.3.4. LEMMA. *If $a\leqslant 0$ and $b\geqslant -1$, then the largest diagonal coefficient of C vanishes. Therefore, $\bigcup_{k\geqslant 1}R_k$ does not intersect the region $a<0$ and $b\geqslant -1$.*

Proof. If a^2+4b is positive, the result follows from Lemma 11.3.3. If a^2+4b is negative, equation (11.3.1)–(11.3.3) gives

$$\begin{aligned}C_{n-k,n-k}&=A_{2,1}A_{1,1}+A_{3,1}A_{2,1}+\sum_{3\leqslant j\leqslant k}(aA_{j,1}+bA_{j-1,1})A_j\\&=b\,C_{n-k+1,n-k+1}+a\sum_{1\leqslant j\leqslant k}A_j^2\,.\end{aligned}$$

Consequently, for $k\geqslant 2$,

$$C_{n-k,n-k}=b^2C_{n-k+2,n-k+2}+a(b+1)\sum_{1\leqslant j\leqslant k-1}A_j^2+aA_k^2\,.$$

Thus, if $b \geqslant -1$ and $a \leqslant 0$,

$$C_{n-k,n-k} \leqslant b^2 C_{n-k+2,n-k+2} \,.$$

Since $C_{n,n} = 0$ and $C_{n-1,n-1} = a < 0$ on the given range, this shows that $C_{n-k,n-k} \leqslant 0$ for all $k \geqslant 0$. ■

To study the domain $a \leqslant 0$, $b \leqslant -1$ and $b < -a^2/4$ is much more complicated. We assume from now on, and until the end of the proof of Theorem 11.3.2, that (a,b) is in this domain. It is then convenient to make a change of parameterization, setting

$$a = -2r\cos\phi \,, \qquad b = -r^2 \,, \qquad r \geqslant 0 \,, \quad 0 \leqslant \phi \leqslant \pi/2 \,.$$

First, this allows us to obtain a closed formula for the diagonal coefficients of C.

13.3.5. LEMMA. *If $a = -2r\cos\phi$ and $b = -r^2$, with r nonnegative and ϕ in $[\,0,\pi/2\,]$, then*

$$\begin{aligned} C_{n-k+1,n-k+1} = r\Big[& -2(1-r^{2k})\cos\phi\sin^2\phi \\ & + r^{2(k-1)}(1-r^2)\sin(k\phi)\Big(\sin\big((k+1)\phi\big) - r^2\sin\big((k-1)\phi\big)\Big)\Big]\Big/ \\ & \qquad (1-r^2)\big((1-r^2)^2 + 4r^2\sin^2\phi\big)\sin^2\phi \,. \end{aligned}$$

Proof. Write $u = re^{i\theta}$ and $v = re^{-i\theta}$ for the roots of the characteristic equation $x^2 - ax - b = 0$. Setting $\phi = \pi - \theta$, we obtain

$$\begin{aligned} a &= u + v = 2r\cos\theta = -2r\cos\phi \\ b &= -uv = -r^2 \,. \end{aligned}$$

This is the origin of the parameterization. Equations (11.3.3) and (11.3.4) give for $k \geqslant 2$,

$$\begin{aligned} C_{n-k+1,n-k+1} &= \sum_{1\leqslant j\leqslant k-1} \frac{u^{2j+1} - (u+v)(uv)^j + v^{2j+1}}{(u-v)^2} \\ &= \frac{1}{(u-v)^2}\Big(u^3 \frac{1-u^{2(k-1)}}{1-u^2} - (u+v)uv\frac{1-(uv)^{k-1}}{1-uv} \\ &\qquad\qquad + v^3\frac{1-v^{2(k-1)}}{1-v^2}\Big) \end{aligned}$$

In this last expression, a part is independent of k. It is the ratio of

$$u^3(1-uv)(1-v^2)-(u+v)uv(1-u^2)(1-v^2)+v^3(1-u^2)(1-uv) \\ = (u-v)^2(u+v) = -4r^2\sin^2\theta\, 2r\cos\theta$$

and

$$\begin{aligned}(u-v)^2&(1-u^2)(1-v^2)(1-uv)\\ &= (u-v)^2(1-uv)\big(-(u+v)^2+(1+uv)^2\big)\\ &= -(2r\sin\theta)^2(1-r^2)\big((1+r^2)^2-4r^2\cos^2\theta\big)\\ &= -4r^2\sin^2\theta\,(1-r^2)\big((1-r^2)^2+4r^2\sin^2\theta\big)\,. \qquad (11.3.5)\end{aligned}$$

For the part dependent on k, we reduce it to the same denominator, (11.3.5), and obtain the numerator

$$\begin{aligned}(1-uv)&\big(-u^{2k+1}-v^{2k+1}+u^2v^2(u^{2k-1}+v^{2k-1})\big)\\ &\qquad\qquad +(u+v)(uv)^k(1-u^2)(1-v^2)\\ &= -(1-r^2)\big(2r^{2k+1}\cos(2k+1)\theta - r^4 2r^{2k-1}\cos(2k-1)\theta\big)\\ &\qquad\qquad +2r\cos\theta\, r^{2k}\big((1-r^2)+4r^2\sin^2\theta\big)\,.\end{aligned}$$

Adding the part independent of k and that dependent of k, we obtain the numerator

$$\begin{aligned}-8(r^3-r^{2k+3})\sin^2\theta\cos\theta-2r^{2k+1}(1-r^2)\big(\cos(2k+1)\theta-r^2\cos(2k-1)\theta\\ -(1-r^2)\cos\theta\big)\\ = -8(r^3-r^{2k+3})\sin^2\theta\cos\theta+4r^{2k+1}(1-r^2)\sin k\theta\big(\sin(k+1)\theta\\ -r^2\sin^2(k-1)\theta\big)\,.\end{aligned}$$

Consequently,

$$\begin{aligned}C_{n-k+1,n-k+1} = r\Big[&2(1-r^{2k})\sin^2\theta\cos\theta\\ &-r^{2(k-1)}(1-r^2)\sin k\theta\big(\sin(k+1)\theta-r^2\sin(k-1)\theta\big)\Big]\Big/\\ &(1-r^2)\big((1-r^2)^2+4r^2\sin^2\theta\big)\sin^2\theta\,.\end{aligned}$$

The change of angle $\theta = \pi-\phi$ gives the result. ∎

In the domain $a \leqslant 0$ and $b < -1$ with $b^2 < -a^2/4$, that is $r > 1$ in our (r,ϕ)-parameterization, Lemma 11.3.5 shows that the sign of

$C_{n-k+1,n-k+1}$ is that of minus its numerator. Thus, it has the same sign as

$$2(1-r^{2k})\cos\phi\sin^2\phi \\ - r^{2(k-1)}(1-r^2)\sin k\phi\big(\sin(k+1)\phi - r^2\sin(k-1)\phi\big)\,.$$

In other words, the sign of $C_{n-k+1,n-k+1}$ is that of

$$2\cos\phi\sin^2\phi + r^{2(k-1)}g_k(r^2) \tag{11.3.6}$$

where

$$\begin{aligned} g_k(s) = &-\sin(k\phi)(\sin k\phi\cos\phi + \cos k\phi\sin\phi) \\ &+ s(-2\cos\phi\sin^2\phi + 2\sin^2 k\phi\cos\phi) \\ &- s^2\sin k\phi(\sin k\phi\cos\phi - \cos k\phi\sin\phi)\,. \end{aligned}$$

We can now explain how to conclude the proof. As k tends to infinity, the leading term in (11.3.6) is $r^{2(k-1)}g(r^2)$, since we assume $r>1$. Thus, for large k the diagonal coefficient $C_{n-k+1,n-k+1}$ is positive whenever $g_k(r^2)$ is such. Thus, our goal is to determine for which values of (s,ϕ) we can have $g_k(s)$ positive for infinitely many k's.

The trick is to understand that when ϕ is an irrational multiple of 2π, the sequence $\sin k\phi$ fills $[-1,1]$. Thus, we can consider $\sin k\phi$ almost as a free parameter, on which we can optimize. This leads us to define the function

$$\begin{aligned} h(\theta) = &-\sin\theta\,(\sin\theta\cos\phi + \cos\theta\sin\phi) \\ &+ s(-2\cos\phi\sin^2\phi + 2\sin^2\theta\cos\phi) \\ &- s^2\sin\theta(\sin\theta\cos\phi - \cos\theta\sin\phi)\,. \end{aligned}$$

Formally, this function is obtained by substituting $k\phi$ for θ in the expression of g_k. It is convenient to define

$$A = (s-1)^2\cos\phi\,,\quad B = (s^2-1)\sin\phi\,,\quad \text{and}\quad C = 2s\sin^2\phi\cos\phi\,.$$

Since we implicitly made the change of variable $s=r^2$ after (11.3.6), the numbers A, B and C are all positive — recall $0<\phi<\pi/2$ and $r>1$.

The following result will be instrumental.

11.3.6. LEMMA. *The function $h(\theta)$ is maximal at a point θ^*, unique modulo π, and defined by*

$$\cos(2\theta^*) = A/\sqrt{A^2+B^2}\,, \qquad \sin(2\theta^*) = B/\sqrt{A^2+B^2}\,.$$

Its maximum value is

$$h(\theta^*) = -\frac{A}{2} + \frac{\sqrt{A^2+B^2}}{2} - C\,.$$

Proof. We rewrite the function $h(\cdot)$ as

$$\begin{aligned} h(\theta) &= -A\sin^2\theta + B\sin\theta\cos\theta - C \\ &= -A\frac{1-\cos 2\theta}{2} + B\frac{\sin 2\theta}{2} - C\,. \end{aligned}$$

Differentiating with respect to θ, we see that when h is maximum,

$$0 = h(\theta^*) = -A\sin 2\theta^* + B\cos 2\theta^*\,.$$

Since neither A nor B vanish, this gives us $\tan\theta^* = B/A$. Consequently, there exists ϵ_1, ϵ_2 equal to either -1 or $+1$, such that

$$\cos 2\theta^* = \epsilon_1 A/\sqrt{A^2+B^2} \qquad \text{and} \qquad \sin 2\theta^* = \epsilon_2 B/\sqrt{A^2+B^2}\,.$$

At such point, the value of $h(\cdot)$ is

$$h(\theta^*) = -\frac{A}{2} + \frac{\epsilon_1 A^2 + \epsilon_2 B^2}{2\sqrt{A^2+B^2}} - C\,.$$

It is maximum when $\epsilon_1 = \epsilon_2 = 1$. This determines 2θ modulo 2π, and therefore, θ modulo π. ■

We can now determine when the maximum of $h(\cdot)$ is positive.

11.3.7. LEMMA. *In the domain $a < 0$ and $b < -1$ with $b < -a^2/4$, the function $h(\cdot)$ has a positive supremum if and only if $b < a - 1$.*

Proof. Using Lemma 11.3.6, the positivity of the supremum of $h(\cdot)$ is equivalent to

$$\sqrt{A^2+B^2} > A + 2C\,.$$

Since A, B and C are positive, this is equivalent to $B^2 > 4C^2 + 4AC$. Plugging the expression for A, B and C into this last inequality, we obtain

$$(s^2-1)^2 > 16s^2(1-\cos^2\phi)\cos^2\phi + 8s(s-1)^2\cos^2\phi\,.$$

Setting $c = \cos^2\phi$, we obtain a quadratic inequality

$$16s^2c^2 - 8s(s^2+1)c + (s^2-1)^2 > 0\,. \qquad (11.3.7)$$

The quadratic function of c involved has two positive roots,

$$c_- = \frac{(s-1)^2}{4s} \qquad \text{and} \qquad c_+ = \frac{(s+1)^2}{4s}\,.$$

Since $c_+ > 1$ (recall $s = r^2 > 0$), inequality (11.3.7) is equivalent to

$$\cos^2\phi < c_-\,.$$

Going back to the parameterization $a = -2r\cos\phi$ and $b = -r^2$, that is $a^2 = 4s\cos^2\phi$ and $b = -s$, we rewrite the above inequality. After a simplification by $4s$, it gives

$$a^2 < (-b-1)^2\,.$$

Since $b < -1$ and $a < 0$, it is equivalent to $b < a - 1$. ∎

We can now state our final lemma.

11.3.8. LEMMA. *Assume that $a < 0$ and $b < \min(-a^2/4, -1)$. If $b < a-1$ and ϕ is an irrational multiple of 2π, then $\limsup_{k\to\infty} g_k(r^2) > 0$. On the other hand, if $b > a-1$, then there exists a positive ϵ such that $g_k(r^2) \leqslant -\epsilon$ for all $k \geqslant 1$.*

Proof. Assume $b < a - 1$. Combining Lemmas 11.3.7 and 11.3.8, let ϵ be a positive number such that $h(\theta)$ is positive on an ϵ-neighborhood of θ^*. If ϕ is an irrational multiple of π, the sequence $k\phi$ intersect $[\,\theta^*-\epsilon, \theta^*+\epsilon\,] + 2\pi\mathbb{Z}$ infinitely often; this follows from Kronecker's approximation theorem in number theory — see, e.g., Hlawka, Schuißengeier and Taschner (1986). Consequently,

$$\limsup_{k\to\infty} g_k(r^2) = h(\theta^*) > 0\,.$$

If $b > a-1$, then $h(\cdot)$ is a negative function. Since $g_k(r^2) = h(k\phi) \leqslant h(\theta^*)$, the result follows. ∎

To conclude the proof of Theorem 11.3.2, we still assume $a < 0$ and $b < \min(-a^2/4, -1)$. If $b < a - 1$, Lemma 11.3.8 shows that whenever ϕ is an irrational multiple of π, the limit superior of (11.3.6) is $+\infty$. For such values of ϕ, the pair (a, b) is covered by infinitely many regions R_k. Thus in the range $b < a - 1$, the only regions not eventually covered by $\bigcup_{k\geqslant 1} R_k$ are those for which ϕ is a rational multiple of π. After the change of parameterization, the complement of this potentially uncovered set is dense in $b < a - 1$.

If $b > a - 1$, then (11.3.6) is less than

$$2\cos\phi\sin^2\phi + r^{2(k-1)}h(\theta^*)\,, \tag{11.3.8}$$

which tends to $-\infty$ as k tends to infinity. Therefore, the pair (a, b) can be covered by at most a finite number of regions R_k. This concludes the proof of Theorem 11.3.2. ∎

Notice that we proved much more than the statement of Theorem 11.3.2. When $a < 0$ and $b > -a^2/4$, Lemma 11.3.3 shows that no region R_k covers (a, b).

The proof of Lemma 11.3.4 also contains useful information. If $a < 0$ and $b > -a^2/4$, the sequence $k \mapsto C_{n-k,n-k}$ is decreasing. Thus, the constant $c(a, b)$ in Theorem 11.3.1 and given by Theorem 8.3.1 is

$$c(a,b) = K_{s,\alpha}^2\alpha^\alpha \sum_{1\leqslant j\leqslant n} |C_{n,j} + C_{j,n}|^\alpha\,.$$

This expression simplifies further if one notices that

$$C_{n,j} = \sum_{2\leqslant i\leqslant n} A_{i,n}A_{i-1,j} = A_{n,n}A_{n-1,j} = A_{n-1,j}\,.$$

Thus, $C_{n,j} = A_{n-j-2,1}$ for $1 \leqslant j \leqslant n-1$. Moreover

$$C_{j,n} = 0 \qquad \text{for all } 1 \leqslant j \leqslant n\,.$$

Consequently,

$$c(a,b) = K_{s,\alpha}^2\alpha^\alpha \sum_{1\leqslant j\leqslant n-1} |A_{n-1,j}|^\alpha = K_{s,\alpha}^2\alpha^\alpha \sum_{1\leqslant j\leqslant n-1} |A_{n-j,1}|^\alpha\,.$$

In the range $a < 0$ and $a - 1 < b < -a^2/4$, the proof of Lemma 11.3.8 shows that at most a finite number of regions R_k cover (a,b). Notice that the bound (11.3.8) shows that the number of such covering regions is at most the largest k for which (11.3.8) is positive. Ultimately, this gives an inequality in k, a and b. Potentially, this could be used if someone were interested in proving some result for particular values of a and b. However, the pictures above suggest that no region R_k covers such pair a (a,b); but I don't know how to prove it.

The proof of Lemma 11.3.9 involves a number theoretic argument which does not say what happens when ϕ is a rational multiple of π. The pictures of the regions R_i below leave the possibility that some exceptional parabola $a = -r\cos\phi$, $b = -r^2$ with $\phi \in 2\pi\mathbb{Q}$ are left uncovered. Equation (11.3.6) shows that no parabola is left completely uncovered. Indeed, as r tends to infinity, the sign of $C_{n-k+1,n-k+1}$ is that of $\sin k\phi \sin(k-1)\phi$. We claim that the sequence $\sin k\phi \sin(k-1)\phi$ contains infinitely many positive values whenever ϕ is in $(0,\pi/2)$. Indeed, if ϕ is in $(0,\pi/2)$, then $k\phi$ and $(k-1)\phi$ are less than $\pi/2$ apart. When ϕ is a rational multiple of π the sequence $(k\phi)_{k\geqslant 1}$ is periodic modulo 2π. Consequently, for some k, both $k\phi$ and $(k-1)\phi$ are in $(0,\pi)$ modulo 2π. For this specific k we have $\sin k\phi \sin(k-1)\phi > 0$.

I conjecture that Theorem 11.3.2 is sharp, meaning that the region described in (a,b) coincides with $\bigcup_{k\geqslant 1} R_k$.

Combined with our description of $R_1 = \{\,(a,b) : a > 0\,\}$, Lemmas 11.3.4 and 11.3.5 allow us to describe completely what happens in the stability region.

11.3.10. THEOREM. *For the second order autoregressive process, $X_n = aX_{n-1} + bX_{n-2} + \epsilon_n$, assume that the roots of the characteristic equation $x^2 - ax - b = 0$ are inside the unit disk. Then the tail of $n\gamma_n(1)$ has the form*

$$P\{\, n\gamma_n(1) \geqslant t \,\} \sim \begin{cases} c(a,b)t^{-\alpha/2} & \text{if } a > 0 \\ c(a,b)t^{-\alpha}\log t & \text{if } a < 0 \end{cases} \qquad \text{as } t \to \infty\,.$$

In particular, in the stability region, the tail of $n\gamma_n(1)$ behaves like $t^{-\alpha/2}$ if $a > 0$ and like $t^{-\alpha}\log t$ if $a < 0$. In some sense, this generalizes Theorem 11.2.1 to autoregressive processes of order 2. If $b = 0$ we recover Theorem 11.2.1.

Notes

A basic reference on the classical theory and applications of time series is Brockwell and Davis (1987).

For linear processes with heavy tailed errors, Davis and Resnick (1986) developed the asymptotic theory as the number of observations goes to infinity. A slightly different perspective, first letting n tend to infinity, and then looking at the tail of the limiting distribution has been investigated in a series of papers by Mijnheer (1997a, b, c).

When the errors ϵ_i's have a nearly symmetric distribution and enough moments, it is tempting to work as if they were from a Weibull-like distribution, or even from a normal one. In the later case, the autocovariances are weighted sums of chi-square random variables. I don't know how to assess the accuracy of such an approximation.

I somewhat believe that a proper understanding of the regions R_i in general requires adding some algebraic geometric tools. Cox, Little and O'Shea (1992, 1998) may be a good starting point for statisticians interested in pursuing this research path. I don't know what is the analogue of Theorem 11.3.2 for autoregressive models of order larger than or equal to 3.

One may think that our choice of heavy tailed errors is the origin of the complicated tail behavior of the autocovariances. If the ϵ_i's have a spherical distribution, we can use the classical Gaussian trick of diagonalizing the matrix of the quadratic form. When looking at the tail behavior, one is then led to the nontrivial question of relating the dimension of the largest eigensubspace of the matrix C to the parameter θ of the autoregressive model. A plot of the spectral gap for the matrix C as a function of a and b — thus, for autoregressive of order 2 — suggests an incredibly complex behavior of the tail of the the aucovariance, even with Gaussian errors.

12. Suprema of some stochastic processes

In chapter 9, we studied how the tail of the distribution of the supremum of a random linear form is related to integration over some asymptotic sets. The goal in this chapter is to go a little further, considering some examples which are of pedagogical interest.

12.1. Maxima of processes and maxima of their variances.

Consider a centered Gaussian process $X(m)$ indexed by some abstract set M. As in chapter 9, write

$$X(M) = \sup\{\, X(m) \,:\, m \in M\,\}$$

for its supremum. Define

$$\sigma^2(M) = \sup\{\, \mathrm{Var} X(m) \,:\, m \in M\,\}$$

to be the supremum of its variance. A famous result of Fernique (1970), Landau and Shepp (1970) asserts that whenever $\sigma^2(M)$ is finite,

$$\lim_{t\to\infty} t^{-2} \log P\{\, X(M) \geqslant t\,\} = -\frac{1}{2\sigma^2(M)}\,. \tag{12.1.1}$$

This result is often interpreted in saying that, in the logarithmic scale, the tail of the supremum of the process is driven by the points of largest variance. A heuristic argument of why this should be the case is that when the variance is large, the process tends to fluctuate more. And so, we should expect its supremum, when large, to be near such a point. As it is, this heuristic argument could be applied to any process. The aim of this section is to show that this heuristic is wrong.

We are going to construct some processes whose supremum tail is driven by the points of smallest variance. This is an application of ideas developed in chapter 9. The key is to understand why the maximum variance appears in the Gaussian case. Proposition

9.2.1 tells us that this happens because $I_\bullet$ is proportional to $1/|x|^2$ in the Gaussian case, as x tends to infinity. That is $I_\bullet$ is inversely proportional to the Euclidean norm — a very specific feature. And the Euclidean norm is a monotone function of the variance. Thus, to build a counterexample to the heuristic, we need to have a set M and a function I such that $I_\bullet$ is minimal on M at points of minimal Euclidean norm.

Let α be in $(1,2)$ and $\theta = \big(\Gamma(1/\alpha)/\Gamma(3/\alpha)\big)^{\alpha/2}$. The function

$$f_\alpha(x) = \frac{\alpha}{2\sqrt{\Gamma(1/\alpha)\Gamma(3/\alpha)}} e^{-\theta|x|^\alpha}\,, \qquad x \in \mathbb{R}\,,$$

defines a density. It has zero expectation, and unit variance. For any positive β, define the unit sphere for the ℓ_β-norm

$$S_d^{(\beta)} = \{\, x \in \mathbb{R}^d \,:\, |x|_\beta = 1 \,\}\,.$$

If $\beta > 1$, define also

$$M_\beta = \{ \sum_{1\leqslant i\leqslant d} \mathrm{sign}(m_i)|m_i|^{\beta-1} e_i \,:\, m \in S_d^{(\beta)} \,\}\,.$$

12.1.1 THEOREM. *Let $X = (X_1, \ldots, X_d)$ be a random vector in $\mathbb{R}^d$ with independent components, all distributed with density f_α. If $1 < \alpha < \beta < 2$, then*

$$\lim_{t\to\infty} t^{-\alpha} \log P\{\, X(M_\beta) \geqslant t \,\} = -\frac{1}{\inf_{m\in M_\beta} \mathrm{Var} X(m)}\,.$$

Proof. For x in $\mathbb{R}^d$, let $I(x) = \sum_{1\leqslant i\leqslant d} \log f(x_i)$. For some constant c,

$$I(x) = \theta \sum_{1\leqslant i\leqslant d} |x_i|^\alpha + c\,.$$

This is a strictly convex function since $\alpha > 1$. It satisfies (9.2.1). Moreover, Lemma 9.1.9 gives

$$I_\bullet(x) = c + |x|^{-\alpha}_{\alpha/(\alpha-1)}\,.$$

Consequently, as t tends to infinity,

$$\begin{aligned} I_\bullet(M_\beta/t) &\sim t^{-\alpha} \Big/ \sup\Big\{ \sum_{1\leqslant i\leqslant d} |x_i|^{\alpha/(\alpha-1)} \,:\, x \in M_\beta \Big\}^{\alpha-1} \\ &= t^{-\alpha} \Big/ \sup\Big\{ \sum_{1\leqslant i\leqslant d} |m_i|^{\frac{(\beta-1)\alpha}{\alpha-1}} \,:\, m \in S_d^{(\beta)} \Big\}^{\alpha-1}\,. \end{aligned}$$

If m belongs to $S_d^{(\beta)}$, then each component $|m_i|$ is less than or equal to 1. Then, the inequality $(\beta-1)/\beta > (\alpha-1)/\alpha$ gives

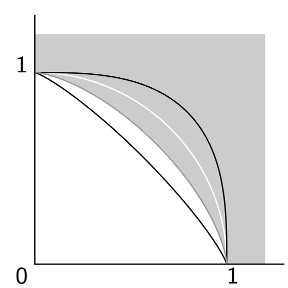
$$\Big(\sum_{1\leqslant i\leqslant d}|m_i|^{\frac{(\beta-1)\alpha}{\alpha-1}}\Big)^{\alpha-1} \leqslant \Big(\sum_{1\leqslant i\leqslant d}|m_i|^{\beta}\Big)^{(1/\beta)\alpha\beta} \leqslant 1\,,$$

with equality if and only if m has exactly one coordinate equal to 1 or -1. Thus Proposition 9.2.1 yields

$$\lim_{t\to\infty} t^{-\alpha}\log P\{\,X(M)\geqslant t\,\} = -1\,.$$

Next the variance of $X(m)$ is $|m|^2$. But if m belongs to M_β, the inequality $|m|^2 \geqslant 1$ holds, with equality if and only if one of the components of m is 1 or -1, and all the others are zero. This concludes the proof. ∎

The proof we gave is reasonably short, but a bit mysterious. For $d=2$, the following picture makes the result obvious if one keeps in mind Laplace's method. It represents the upper right quadrant of the plane, and the various sets involved, for $\alpha=1.2$ and $\beta=1.5$. The set on which the integration is performed is shaded gray. Its boundary is the polar reciprocal of M_β. The set M_β is the black line inside the shaded area. The Euclidean unit sphere is the white line inside the shaded area. The last black line is the level set of the density $f_\alpha(x)f_\alpha(y)$.



Notice that for $d=2$ we have a process indexed by the one dimensional set M_β. For $d>2$ we have a random field. More importantly, Theorem 12.1.1 may be a prototype for a misleading statement! The proof shows that the left hand side of the statement has no intrinsic connection with the right hand side! The proof shows that indeed the variance appears by a pure coincidence. In my opinion, the Gaussian case is not any different, as Proposition 9.2.1 shows.

12.2. Asymptotic expansions for the tail of the supremum of Gaussian processes can be arbitrarily bad.

For many reasons, both theoretical and applied, there has been a large literature devoted to the approximation of the tail probability of the supremum of Gaussian processes. In view of (12.1.1), it is quite natural to search for an approximation of the form

$$P\{\, X(M) \geqslant t\,\} = t^{\alpha} \exp\Big(\frac{-t^2}{2\sigma^2(M)}\Big)\big(P_n(1/t)+o(t^{-n})\big) \qquad \text{as } t\to\infty \tag{12.2.1}$$

where P_n is a polynomial of degree n. The hope of course is that for moderate t's, this expansion provides an accurate approximation.

The aim of this section is to provide an example where such expansion holds, but, no matter what, provides a poor approximation for fixed t. It is essential to remember that an asymptotic expansion like (11.2.1), if it exists, is unique — see, e.g., Olver, 1974, §1.7. In particular, it does not depend at all on which method is used to derive it. Thus, the failure we want to describe is not that of a particular method. The one proposed in these notes as well as any other fails, and there is no way arround if one sticks to approximations of the form (12.2.1). The basic idea in this section is to mimic what happened for autoregressive models. Our approximation was not so good when we appealed to Theorem 8.3.1, because it was actually quite likely that the largest random variable was not an ϵ_i for which $C_{i,i}=0$; even if one conditions by the appearance of a large deviation, it is quite likely that it is caused by a large ϵ_i for which $C_{i,i}=0$.

To build our example we will first give it in a geometric form. We will discuss afterwards some of its features.

Consider the convext set in $\mathbb{R}^d$,

$$C = \{\, p \,:\, \langle p, e_1\rangle \leqslant 1\,\} \bigcap \{\, p \,:\, |p| \leqslant 1+\epsilon\,\}\,.$$

Denote by M the polar reciprocal of ∂C. Let X be a random vector in $\mathbb{R}^d$, having a centered normal distribution, with independent components. Then

$$X(M) \geqslant t \text{ if and only if } X \notin tC\,.$$

Define the polynomials

$$P_n(u) = \sum_{0\leqslant k\leqslant n} (-1)^k \frac{(2k)!}{2^{2k}k!} u^k\,.$$

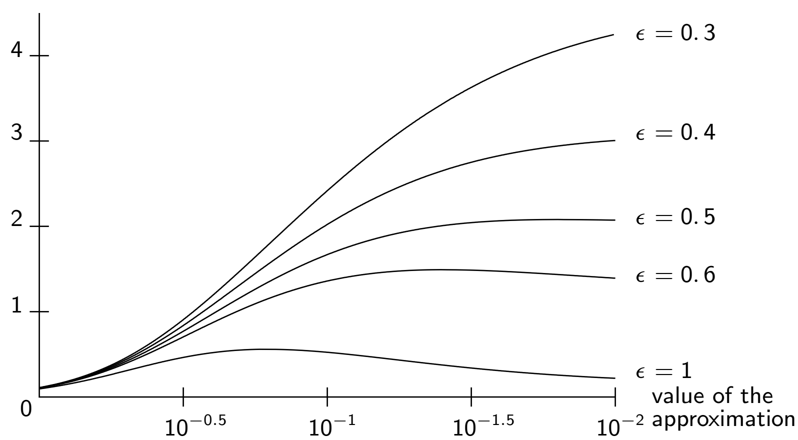
12.2.1 PROPOSITION. *The tail expansion*

$$P\{\,X(M)\geqslant t\,\}=\frac{e^{-t^2/2}}{\sqrt{2\pi}t}\big(P_n(1/t)+o(t^{-n})\big)\qquad \textit{as } t\to\infty$$

holds. However, for any $t\geqslant 0$,

$$\frac{\sqrt{2\pi}te^{t^2/2}}{P_n(1/t)}P\{\,X(M)\geqslant t\,\}\geqslant 2\sqrt{2\pi}e^{-t^2(\epsilon+(\epsilon^2/2))}t^{d-(1/2)}\,.$$

Before proving this result, let us see why this provides the proper example. The first statment lets us hope that $e^{-t^2/2}P_n(1/t)/\sqrt{2\pi}t$ is a good approximation of the tail probability of the supremum. The second statment asserts that it is not the case if ϵ is small and t is large, but not too large. The following plot shows the lower bound for $d=2$, $n=2$ and various values of ϵ, as a function of the approximation. For instance, if $\epsilon=1$, the lower bound is less than 1, which means that the approximation may underestimate. A more interesting value is for $\epsilon=0.5$; when the approximation is about $10^{-1.5}\approx 3\%$, the lower bound is about 2. Thus, the approximation underestimates the correct probability by a factor at least 2. One should keep in mind that for a typical statistical application, we are interested in t's such that $P\{\,X(M)\geqslant t\,\}$ is between 10^{-1} and 10^{-2}.



We will comment further on this example after we prove our statement.

Proof of Proposition 12.2.1. We first start with the obvious bound

$$
\begin{aligned}
P\{\, X(M) \geqslant t \,\} &= P\{\, X_1 \geqslant t \text{ or } |X| \geqslant t(1+\epsilon) \,\} \\
&\begin{cases} \geqslant P\{\, X_1 \geqslant t \,\} \\ \leqslant P\{\, X_1 \geqslant t \,\} + P\{\, |X| \geqslant t(1+\epsilon) \,\} \,. \end{cases}
\end{aligned}
\tag{12.2.1}
$$

The crude logarithmic estimate of Proposition 2.1 shows that

$$
P\{\, |X| \geqslant t(1+\epsilon) \,\} = O(e^{-t^2(1+\epsilon)^2/2} t^{2d}) \qquad \text{as } t \to \infty \,.
$$

On the other hand, the standard asymptotic expansion for the complementary error function (see, e.g., Olver, 1974, §3.1.1) yields

$$
P\{\, X_1 \geqslant t \,\} = \frac{e^{-t^2/2}}{\sqrt{2\pi} t} \big(P_n(1/t) + o(t^{-n}) \big) \qquad \text{as } t \to \infty \,.
$$

Thus (12.2.1) provides the asymptotic expansion in the first assertion of Proposition 12.2.1.

We now use another lower bound, namely

$$
P\{\, X(M) \geqslant t \,\} \geqslant P\{\, |X| \geqslant t(1+\epsilon) \,\} \,,
$$

which follows from the equality in (12.2.1). Since $|X|^2$ has a chi-square distribution with d degrees of freedom, an integration by parts yields

$$
\begin{aligned}
P\{\, |X| \geqslant t(1+\epsilon) \,\} &= \int_{t^2(1+\epsilon)^2}^{\infty} \frac{x^{\frac{d}{2}-1} e^{-x/2}}{2^{d/2}\Gamma(d/2)} dx \\
&\geqslant 2 t^{d-\frac{1}{2}} (1+\epsilon)^{d-\frac{1}{2}} e^{-t^2(1+\epsilon)^2/2} \,.
\end{aligned}
$$

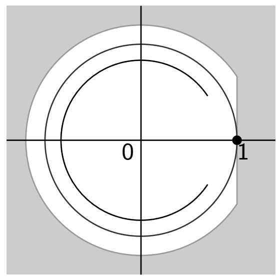
For any $u \geqslant 0$, the bound $P_n(u) \leqslant 1$ holds; this comes from the fact that the asymptotic expansion for the error function is obtained by integrating by parts, and the integrations lead to an alternating series — see, e.g., Olver, 1974, §3.1. Thus the second statement of Proposition 12.2.1 follows. ■

When $d = 2$, we can make a very explicit construction of the process. The polar reciprocal M is just a piece of circle and a point,

$$
M = \big\{\, (x,y) \in \mathbb{R}^2 \,:\, x^2 + y^2 = (1+\epsilon)^{-2} \,; x \leqslant (1+\epsilon)^{-2} \,\big\} \cup \{\, (1,0) \,\} \,.
$$

This can be seen by a formal proof, but it is obvious from the following picture. The domain on which we integrate is shaded. The unit

sphere, or equivalently the level set of the Gaussian measure, is the dark sphere. The set M is an open arc, in black as well, with the point e_1 marked.



We choose M to be reduced — see definition in section 9.1. We could as well index the process by a larger set, such as

$$(1+\epsilon)^{-1}S_1 \cup \{\,(1,0)\,\}$$

or even

$$(1+\epsilon)^{-1}S_1 \cup [\,(1+\epsilon)^{-1}e_1, e_1\,]\,.$$

This last set can be parameterized as follows. Let

$$f(t) = \begin{cases} \dfrac{1}{1+\epsilon}\big(\cos(2\pi t), \sin(2\pi t)\big) & \text{if } 0 \leqslant t \leqslant 1 \\ \Big(\dfrac{\epsilon t + 1 - \epsilon}{1+\epsilon}, 0\Big) & \text{if } 1 \leqslant t \leqslant 2. \end{cases}$$

The corresponding Gaussian process is

$$X(t) = \begin{cases} \dfrac{1}{1+\epsilon}\big(X_1\cos(2\pi t) + X_2\sin(2\pi t)\big) & \text{if } 0 \leqslant t \leqslant 1 \\ \dfrac{\epsilon t + 1 - \epsilon}{1+\epsilon}X_1 & \text{if } 1 \leqslant t \leqslant 2. \end{cases}$$

Its variance is

$$\mathrm{Var}X(t) = \begin{cases} \dfrac{1}{(1+\epsilon)^2} & \text{if } 0 \leqslant t \leqslant 1 \\ \Big(\dfrac{\epsilon t + 1 - \epsilon}{1+\epsilon}\Big)^2 & \text{if } 1 \leqslant t \leqslant 2. \end{cases}$$

Now one can argue that our example is specific; the maximal variance is achieved when $t = 2$, that is on the boundary of the domain. Well, we can always define

$$Y(t) = \begin{cases} X(t) & \text{if } 0 \leqslant t \leqslant 2 \\ X(4-t) & \text{if } 2 \leqslant t \leqslant 4. \end{cases}$$

This process has maximum variance at $t = 2$. And certainly, one could argue that $\mathrm{Cov}(X(t), X(s)) = 1$ if $1 \leqslant s, t \leqslant 2$, and thus this process is pathological. This argument can be also ruled out by perturbing each coordinate $Y(t)$ by some very tiny multiple of a brownian bridge.

The moral of the story is that the asymptotic expansion should not be worked out blindly. One should certainly make a careful study of the covariance of the process and be very cautious when the variance does not vary much. Notice that the ratio of the maximal variance of the process to the minimal one is $(1+\epsilon)^2$. For $\epsilon = 0.3$, this is 1.69, which is not that small. Going back to the lower bound in Proposition 12.2.1, notice that the polynomial term in the lower bound is of order $t^{d-(1/2)}$. As the dimension d increases, the asymptotic expansion gives a worse approximation. It can be arbitrarily bad by just taking d large enough. Therefore, the constancy of the variance should be measured with respect to the dimension. It is therefore very unclear what happens in large dimensions or even in infinite dimensions. It is also unclear how to assess a priori the approximating quality of an asymptotic expansion in this context.

12.3. Maximum of nonindependent Gaussian random variables.

When one wants to simulate numerically a Gaussian process, there is not much choice other than to discretize it. To what extent can we obtain an approximation of the distribution of the original process by that of the corresponding discretization? There is no claim that this section brings some new result. The one we are going to prove now can be derived from others existing in the literature. But its derivation may be of pedagogical interest.

Having in mind a discretized process, let X be a Gaussian vector in $\mathbb{R}^d$, with mean 0 and definite positive covariance matrix Σ. Let $\sigma^2 = \max_{1\leqslant i\leqslant d} \Sigma_{i,i}$ be a largest diagonal element.

12.3.1. THEOREM. *Let X have a Gaussian distribution, centered, with positive definite covariance matrix Σ. Then*

$$P\{\max_{1\leqslant i\leqslant d} X_i \geqslant t\} \sim \frac{\sigma e^{-t^2/2\sigma^2}}{t\sqrt{2\pi}} \sharp\{\, i : \Sigma_{i,i} = \sigma^2\,\}\,.$$

Proof. We apply Theorem 7.1. Define

$$A_t = tA_1 = t\{\, x \in \mathbb{R}^d : \max_{1\leqslant i\leqslant d} x_i \geqslant 1\,\}\,.$$

Set

$$I(x) = \frac{1}{2} x^{\mathrm{T}} \Sigma^{-1} x \, .$$

This is a convex function, homogenous of degree $\alpha = 2$, and

$$P\{ \max_{1\leqslant i\leqslant d} X_i \geqslant t \} = \frac{1}{(2\pi)^{d/2} (\det \Sigma)^{1/2}} \int_{tA_1} e^{-I(x)} \, \mathrm{d}x \, .$$

To apply Theorem 7.1, we need to minimize I over A_1. Since Σ is symmetric, we can diagonalize it and write $\Sigma = QDQ^{\mathrm{T}}$ with D diagonal and Q orthogonal. The change of variable $x = Q^{\mathrm{T}} y$ shows that

$$\begin{aligned} I(A_1) &= \inf \Bigl\{ \frac{1}{2} x^{\mathrm{T}} \Sigma^{-1} x \; : \; \max_{1\leqslant i\leqslant d} x_i \geqslant 1 \Bigr\} \\ &= \inf \Bigl\{ \frac{1}{2} y^{\mathrm{T}} D y \; : \; \max_{1\leqslant i\leqslant d} \langle y, Q^{\mathrm{T}} e_1 \rangle \geqslant 1 \Bigr\} \, . \end{aligned}$$

Writing the Lagrangian to optimize $y^{\mathrm{T}} D^{-1} y$ subject to the constraint $\langle y, Q^{\mathrm{T}} e_i \rangle = 1$ and optimizing over i, we obtain

$$I(A_1) = \frac{1}{2} \min_{1\leqslant i\leqslant d} \frac{1}{e_i^{\mathrm{T}} D e_i} = \frac{1}{2 \max_{1\leqslant i\leqslant d} \Sigma_{i,i}} = \frac{1}{2\sigma^2} \, .$$

Moreover, $I(A_1)$ is achieved for the points $y = DQ^{\mathrm{T}} e_i / \sigma^2$, or equivalently $x = \Sigma e_i / \sigma^2$. Thus, the dominating manifold for A_1 is

$$\mathcal{D}_{A_1} = \Bigl\{ \frac{\Sigma e_i}{\sigma^2} \; : \; i \text{ such that } \Sigma_{i,i} = \sigma^2 \Bigr\} \, .$$

It is of dimension $k = 0$. Since $\mathrm{D}I = \Sigma^{-1}$, Theorem 7.1 yields

$$P\{ \max_{1\leqslant i\leqslant d} X_i \geqslant t \} \sim \frac{e^{-t^2/2\sigma^2}}{t\sqrt{2\pi} (\det \Sigma)^{1/2}} \sum_{i:\Sigma_{i,i}=\sigma^2} \frac{\sigma^{d+1}}{\bigl(\det G_{A_1}(\Sigma e_i/\sigma^2)\bigr)^{1/2}}$$

as t tends to infinity. We need to calculate G_{A_1}. As mentioned after the statement of Theorem 7.1.1, it is obtained as the compression of the difference of two second fundamental forms. The one for ∂A_1 vanishes since ∂A_1 is locally a flat hyperplane. That for the level set of I is $\mathrm{D}^2 I / |\mathrm{D}I|$. At $\Sigma e_1/\sigma^2$, its value is $\sigma^2 \Sigma^{-1}$. The tangent space at ∂A_1 at this point is

$$\{ \mathrm{D}I(\Sigma e_i/\sigma^2) \}^{\perp} = (e_i/\sigma^2)^{\perp} = e_i^{\perp} \, .$$

Thus, $G_{A_1}(\Sigma e_i/\sigma^2)$ is the expression of $\sigma^2 \Sigma^{-1}$ to $e_i^{\perp}$. Therefore

$$\det G_{A_1}(\Sigma e_i/\sigma^2) = \sigma^{2(d-1)} \det \langle \Sigma^{-1} e_k, e_l \rangle_{\substack{1\leqslant k,l\leqslant d \\ k,l\neq i}} \, .$$

Thus, it is $\sigma^{2(d-1)}$ times the determinant of the (i,i)-cofactor of Σ^{-1}, which is $\Sigma_{i,i}\det\Sigma^{-1}$. Consequently,

$$\frac{\sigma^{d+1}}{(\det\Sigma)^{1/2}\det G_{A_1}(\Sigma e_i/\sigma^2)^{1/2}} = \sigma\,,$$

and this gives the result putting all the estimates together. ∎

The fact we now want to stress is about discretizing Gaussian processes to simulate the distribution of their maximum. Since the maximum of the discretization is less than the maximum of the original process, this can only give a lower bound. On the far tail, Theorem 12.3.1 asserts that this lower bound must be of order $\sigma e^{-t^2/2\sigma^2}/t\sqrt{2\pi}$. This implies two things. First, one should include the points of largest variance in the discretized sequence. This is almost common sense. Second, in theory, the far tail will be well approximated only if that of the original process behaves like $\sigma e^{-t^2/2\sigma^2}/t\sqrt{2\pi}$. These processes have been characterized by Talagrand (1988).

12.4. The truncated Brownian bridge.

The Brownian bridge B on $[\,0,1\,]$ is a centered Gaussian process with covariance

$$\mathrm{E}B(s)B(t) = st - s\wedge t\,.$$

It is a classical result that it admits the Karhunen-Loève expansion

$$B(s) = \frac{\sqrt{2}}{\pi}\sum_{k\geqslant 1} X_k\frac{\sin(k\pi s)}{k}\,, \qquad 0\leqslant s\leqslant 1\,,$$

where the X_k's are independent, normally distributed random variables. It is known — see, e.g., Billingsley, 1968, §11 — that

$$P\{\sup_{0\leqslant s\leqslant 1} B(s)\geqslant t\,\} = e^{-2t^2}\,. \tag{12.4.1}$$

The aim of this section is to obtain an approximation for the tail of the supremum of the truncated series

$$B_d(s) = \frac{\sqrt{2}}{\pi}\sum_{1\leqslant k\leqslant d} X_k\frac{\sin(k\pi s)}{k}\,.$$

This example is quite interesting, because we will see that Theorem 5.1 — or equivalently, Theorem 7.1 — does not apply. However, a

slight change in the arguments will allow us to obtain the desired asymptotic equivalence.

Let us first explain why Theorem 7.1 does not apply, and, in particular, why assumption (7.5) — or assumption (5.3) if one uses Theorem 5.1 — is not satisfied. Recall that $e_1, \ldots, e_d$ denotes the canonical basis in $\mathbb{R}^d$. Define the curve

$$p(s) = \frac{\sqrt{2}}{\pi} \sum_{1 \leqslant k \leqslant d} \frac{\sin(k\pi s)}{k} e_k \,, \qquad 0 \leqslant s \leqslant 1$$

in $\mathbb{R}^d$. Introducing the Gaussian vector $Y = (Y_1, \ldots, Y_d)$, we have $B_d(s) = \langle Y, p(s) \rangle$. Define

$$A_t = \Big\{\, y \in \mathbb{R}^d : \sup_{0 \leqslant s \leqslant 1} \langle y, p(s) \rangle \geqslant t \,\Big\} = tA_1 \,.$$

Let $I(y) = |y|^2/2$. Then

$$P\{ \sup_{0 \leqslant s \leqslant 1} B_d(s) \geqslant t \,\} = \frac{1}{(2\pi)^{d/2}} \int_{A_t} e^{-I(y)} \, \mathrm{d}y \,.$$

To find the dominating manifold, Proposition 9.1.7 combined with Lemma 9.1.9 suggest that we should search for the points s maximizing the variance of $B_d(s)$,

$$\mathrm{Var} B_d(s) = |p(s)|^2 = \frac{2}{\pi^2} \sum_{1 \leqslant k \leqslant d} \frac{\sin^2(k\pi s)}{k^2} \,.$$

Since $\mathrm{Var} B_d(s) = \mathrm{Var} B_d(1-s)$, it suffices to locate the maximum in $[\,0, 1/2\,]$. We differentiate the variance, obtaining

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} \mathrm{Var} B_d(s) &= \frac{2}{\pi^2} \sum_{1 \leqslant k \leqslant d} \frac{2 \sin(k\pi s) \cos(k\pi s)}{k} \\ &= \frac{2}{\pi^2} \sum_{1 \leqslant k \leqslant d} \frac{\sin(2k\pi s)}{k} \,. \end{aligned}$$

It follows from Jackson's (1912) theorem — see, e.g., Andrews, Askey and Roy, 1999, chapter 7 — that $(\mathrm{d}/\mathrm{d}s)\mathrm{Var} B_d(s)$ is positive on $(0, 1/2)$. Thus $B_d(s)$ has a unique point of maximal variance for $s = 1/2$, no matter what d is. The maximal value depends on d. It is

$$\sigma_d^2 = \mathrm{Var} B_d(1/2) = \frac{2}{\pi^2} \sum_{\substack{1 \leqslant k \leqslant d \\ k \text{ odd}}} \frac{1}{k^2} \,.$$

Since the variance is maximum at a unique point, Proposition 9.1.7 suggests that the dominating manifold in our problem is a single point $p^* = p(1/2)/|p(1/2)|^2$. Its dimension is $k = 0$.

We then calculate $I(A_1) = |p^*|^2/2 = 1/(2\sigma_d^2)$. The level surface $\Lambda_{I(A_1)}$ is the sphere of radius $1/\sigma_d$, centered at the origin. Its second fundamental form is $\sigma_d \mathrm{Id}_{d-1}$ at every point.

Let us now calculate the second fundamental form of ∂A_1. To parameterize ∂A_1, we follow the construction in section 9.1, with the simplification given in section 9.2 for the special case of a radial function $I(\cdot)$. The vector $\tau = p'/|p'|$ is a unit tangent vector to the curve $M = p([\,0,1\,])$. As defined in section 9.1, let ν be a unit normal vector to M in $(T_pM + p\mathbb{R}) \ominus T_pM$, that is

$$\nu = \frac{p - \langle p, \tau\rangle \tau}{\sqrt{|p|^2 - \langle p, \tau\rangle^2}} \,.$$

Notice that $|p(s)|^2$ being maximal for $s = 1/2$, the vectors $p(1/2)$ and $\tau(1/2)$ are orthogonal, and $\nu(1/2)$ equals $p^*|p(1/2)|$.

Let $X_1, \ldots, X_{d-1}$ be an orthonormal moving frame in $\mathbb{R}^d \ominus (T_p\mathbb{R} + p\mathbb{R})$. Using the simplification pointed out after (9.2.4),

$$X(s, u_2, \ldots, u_{d-1}) = \frac{\nu(s)}{\langle \nu(s), p(s)\rangle} + \sum_{2\leqslant j\leqslant d-1} u_j X_j \qquad (12.4.1)$$

defines a local parameterization of ∂A_1, or equivalently of ∂C_M, near p^* when s is chosen close to $1/2$ and $(u_2, \ldots, u_{d-1})$ close to 0. To calculate the second fundamental form of ∂A_1, recall that $p(s)$ is an outward normal to ∂A_1 at $X(s, u_2, \ldots, u_{d-1})$ thanks to Lemma 9.1.6. In particular,

$$T_X \partial C_M = \{\, p\,\}^\perp = \mathrm{span}\{\, X_2, \ldots, X_{d-1}\,\} \oplus \big(\mathrm{span}(p,\tau) \ominus p\mathbb{R}\big) \,.$$

We complete $X_2, \ldots, X_{d-1}$ into an orthonormal basis of $T_X\partial C_M$ by adding the vector field

$$e = \frac{\mathrm{Proj}_{p^\perp}\tau}{|\mathrm{Proj}_{p^\perp}\tau|} = \frac{1}{|p|}\frac{\tau|p|^2 - \langle \tau, p\rangle p}{\sqrt{|p|^2 - \langle \tau, p\rangle^2}} \,.$$

We have

$$\mathrm{d}p \cdot X_j = 0 \qquad \text{for } j = 2, \ldots, d-1 \,,$$

expressing the fact that ∂A_1 is a ruled surface with flat generators in the space spanned by $X_2, \ldots, X_{d-1}$. Thus, the second fundamental

form of ∂A_1 vanishes on $X_2, \ldots, X_{d-1}$. Its matrix in the basis $e, X_2, \ldots, X_{d-1}$ is then

$$\begin{pmatrix} \frac{\langle \mathrm{d}p \cdot e, e \rangle}{|p|} & & 0 \\ & 0 & \\ 0 & & \ddots & \\ & & & 0 \end{pmatrix} .$$

It remains to calculate $\langle \mathrm{d}p \cdot e, e \rangle$ on the dominating manifold $p^* = X(1/2, 0, \ldots, 0)$. The vectors $p(1/2)$ and $\tau(1/2)$ being orthogonal, $e(1/2, 0, \ldots, 0)$ and $\tau(1/2)$ are equal. Since τ is a unit tangent vector to the curve $p([\,0,1\,])$, we have

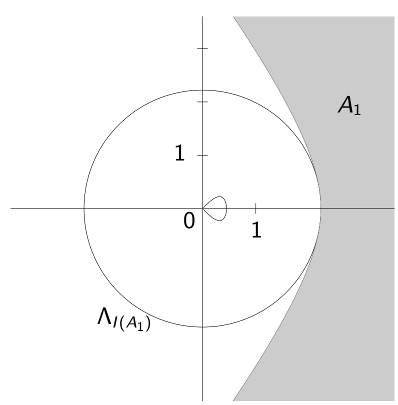
$$\begin{aligned} \langle \mathrm{d}p(1/2) \cdot e(1/2, 0 \ldots, 0), e(1/2, 0, \ldots, 0) \rangle &= \langle \mathrm{d}p \cdot \tau, \tau \rangle (1/2) \\ &= |\tau(1/2)|^2 = 1 \,. \end{aligned}$$

In conclusion, on the dominating manifold M, and in the basis $(e, X_2, \ldots, X_{d-1})$, the fundamental form of ∂A_1 and $\Lambda_{I(A_1)}$ are

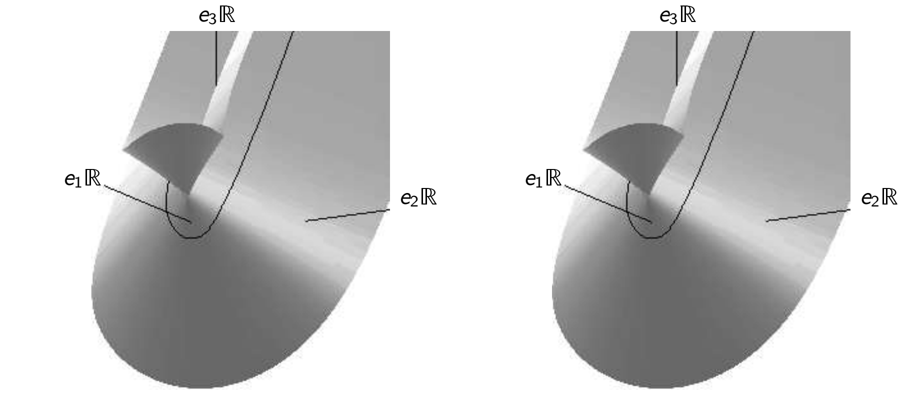
$$\Pi_{\partial A_1} = \begin{pmatrix} \sigma & & 0 \\ & 0 & \\ 0 & & \ddots & \\ & & & 0 \end{pmatrix} \qquad \text{and} \qquad \Pi_{\Lambda_{I(A_1)}} = \begin{pmatrix} \sigma & & 0 \\ & \ddots & \\ 0 & & \sigma \end{pmatrix} .$$

In particular, $\Pi_{\partial A_1} - \Pi_{\Lambda_{I(A_1)}}$ is diagonal. Its upper left entry vanishes. Because the dominating manifold $\mathcal{D}_{\partial C_M}$ is a point, $\Pi_{\partial A_1} - \Pi_{\Lambda_{I(A_1)}} = G_{A_1}$; thus $\det G_{A_1}$ is null, and assumption (7.5) does not hold.

The fact that the fundamental forms have the same upper left entry expresses the fact that along the tangent direction $e(1/2, 0, \ldots, 0)$, the surfaces $\Lambda_{I(A_1)}$ and ∂C_M pull apart very slowly. The following pictures illustrate this fact when $d = 2$ and $d = 3$.

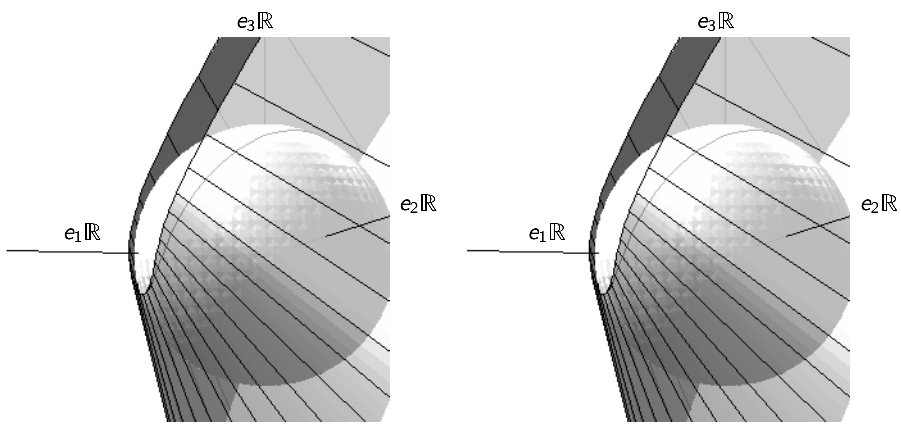


$d = 2$
(the little loop next to the origin is $M = p([0, 1])$.)



$d = 3$

The boundary of A_1 is a ruled surface. This picture shows the ruled surface obtained from the parameterization $X(s, u_2)$, with no constraints on u_2. The actual set A_1 is in the convex hull of the piece containing the origin. The curve drawn on the surface is $X(s, 0) = \nu(s)/\langle \nu(s), p(s) \rangle$.



In this picture, we can see the sphere $\Lambda_{I(A_1)}$, centered and of radius $1/\sigma_3$. The transparent surface is ∂A_1, parametrized by $X(s, u_2)$, but with u_2 constrained to be negative. Its generators are straight lines, parameterized by u_2, shown here for u_2 negative. Thus, the surface ∂A_1 is cut along the curve $X(s, 0)$. The projection of this curve on the sphere $\Lambda_{I(A_1)}$ can be seen. The set ∂A_1 has a unique contact point, p^* with the level set $\Lambda_{I(A_1)}$. But, as it can be seen along the curve $X(s, 0)$, the boundary ∂A_1 pulls away very slowly from $\Lambda_{I(A_1)}$ near p^*.

In the current situation, it is easy to obtain the desired equivalence for the tail of the supremum, because $I(\cdot)$ is homogeneous. If this tail

were behaved like that of the process at its point of maximal variance, we would have a tail equivalence in

$$P\{\, B_d(1/2) \geqslant t \,\} = 1 - \Phi(t/\sigma_d) \sim \frac{1}{\sqrt{2\pi}} \frac{\sigma_d}{t} \exp\Big(-\frac{t^2}{2\sigma_d^2}\Big) .$$

The tail behavior turns out to be much more surprising. Even and odd dimensions d yield different exponents in the polynomial term. The reason is that for even dimensions, the set ∂C_M pulls away more slowly along the direction $e(1/2, 0, \ldots, 0)$. The contact between the two surfaces is of order 1 for odd dimensions, and of order 3 for even ones.

12.4.1. THEOREM. *As t tends to infinity, the ratio*

$$P\{ \sup_{0\leqslant s\leqslant 1} B_d(s) \geqslant t \,\}\big/ \big(1 - \Phi(t/\sigma_d)\big)$$

is equivalent to

(i) $\sqrt{d+1}/\sigma_d$ *if d is odd;*

(ii) $\sqrt{t}\,\Gamma(1/4)\Big(\dfrac{6}{d(d+1)}\Big)^{1/4} \dfrac{\sqrt{d}}{2\pi\sqrt{2}\sigma_d^2}$ *if d is even and $d \neq 1$;*

(iii) 1 *if $d = 1$.*

REMARK. We will see that $\lim_{d\to\infty} \sigma_d^2 = 1/4$. Since $d \mapsto \sigma_d^2$ is increasing, we have $\sigma_d^2 < 1/4$ for any integer d. Thus the constant in statement (i) of Theorem 12.4.1 is always larger than 1. This is in agreement with the failure of assumption 5.3. In the same spirit, the quantity involved in statement (ii) is also larger than 1 when t is large enough.

Proof of Theorem 12.4.1. We first proof statement (iii). Since

$$B_1(s) = \frac{\sqrt{2}}{\pi} X_1 \sin(\pi s) ,$$

its supremum is 0 if X_1 is negative, and $\sqrt{2}X_1/\pi$ otherwise. The result follows.

From now on, assume that $d \geqslant 2$. The change of variable $y = tc$ allows to write

$$\begin{aligned} P\{ \sup_{0\leqslant s\leqslant 1} B_d(s) \geqslant t \,\} &= \frac{1}{(2\pi)^{d/2}} \int_{tA_1} e^{-|y|^2/2} \mathrm{d}y \\ &= \frac{t^d}{(2\pi)^{d/2}} \int_{A_1} e^{-t^2|x|^2/2} \mathrm{d}x \,. \end{aligned}$$

Thus, to prove Theorem 12.4.1, we need to estimate the integral

$$\int_{A_1} e^{-\lambda|x|^2/2}\mathrm{d}x\,,$$

for large values of λ. It is plain from standard results on Laplace's method, or from chapter 7, that we can restrict the integration to the intersection of A_1 with an arbitrary neighborhood of the dominating manifold p^*. Making use of the construction in section 9.1, we can write any point x of A_1 near p^* as

$$x(s,u,v) = \frac{\nu(s)}{\langle \nu(s), p(s)\rangle} + \sum_{2\leqslant j\leqslant d-1} u_j X_j(s) + vp(s)\,, \qquad v \geqslant 0\,,$$

where $u = (u_2, \ldots, u_{d-1})$ is in a neighborhood of the origin. Because $X_2, \ldots, X_{d-1}, \nu$ is an orthonormal basis and p is orthogonal to $X_2, \ldots, X_{d-1}$,

$$|x|^2 = \frac{1}{\langle \nu, p\rangle^2} + |u|^2 + v^2|p|^2 + 2v\,.$$

To calculate the Jacobian of the change of variable $x \leftrightarrow (s,u,v)$, is easy. Write

$$\begin{aligned}
&\frac{\partial x}{\partial u_j} = X_j\,, \qquad \frac{\partial x}{\partial v} = p\,,\\
&\frac{\partial x}{\partial s} = \frac{\partial}{\partial s}\Big(\frac{\nu(s)}{\langle \nu(s), p(s)\rangle}\Big) + \sum_{2\leqslant j\leqslant d-1} u_j \frac{\partial}{\partial s} X_j(s) + vp'(s)\,.
\end{aligned}$$

Introducing $w = p|p'|^2 - \langle p, p'\rangle p'$, we see that $\nu/\langle \nu, p\rangle = w/\langle w, p\rangle$. Thus,

$$\Big(\frac{\nu}{\langle \nu, p\rangle}\Big)' = \frac{w'}{\langle w, p\rangle} - w\frac{\langle w', p\rangle + \langle w, p'\rangle}{\langle w, p\rangle^2}\,.$$

It is straightforward to calculate

$$w' = 2p\langle p', p''\rangle - \langle p, p''\rangle p' - \langle p, p'\rangle p''\,.$$

Since $|p(s)|^2$ is maximal at $s = 1/2$, the tangent vector $p'(1/2)$ is orthogonal to $p(1/2)$. The specific form of $p(\cdot)$ in this problem yields that for any j, the derivative $\mathrm{d}^j p/\mathrm{d}s^j$ at $s = 1/2$ involves only the vectors e_k with k odd if j is, and k even otherwise. Consequently, $p'(1/2)$ and $p''(1/2)$ are orthogonal, and

$$\begin{aligned}
&w(1/2) = p|p'|^2(1/2)\,, \qquad w'(1/2) = -\langle p, p''\rangle p'(1/2)\\
&\frac{\partial}{\partial s}\Big(\frac{\nu(s)}{\langle \nu(s), p(s)\rangle}\Big)(1/2) = -\frac{\langle p, p''\rangle p'}{|p|^2|p'|^2}(1/2)\,.
\end{aligned}$$

Thus, in the orthonormal basis $\tau, X_2, \ldots, X_{d-1}$, the Jacobian matrix is

$$\begin{pmatrix} -\frac{\langle p, p''\rangle}{|p|^2|p'|}(1/2) & & 0 \\ & 1 & \\ 0 & & \ddots \\ & & & 1 \end{pmatrix} + O(|x-p^*|) \qquad \text{as } x \to p^* \,.$$

Now, let ϵ be an arbitrary positive real number. In what follows, η denotes a positive real number, which we will choose as small as needed. Denote by $D(\eta)$ the domain

$$D(\eta) = \{ x(s,u,v) \in \mathbb{R}^d \,:\, |s-1/2| \leqslant \eta \,,\, |u| \leqslant \eta \,,\, 0 \leqslant v \leqslant \eta \,\} \,.$$

From our evaluation of the Jacobian, if η is small enough

$$\begin{aligned} &\int_{A_1 \cap D(\eta)} e^{-\lambda |x|^2/2} \mathrm{d}x \\ &\quad\leqslant (1+\epsilon) \Big| \frac{\langle p, p''\rangle}{|p|^2|p'|}(1/2) \Big| \times \\ &\qquad \int_{D(\eta)} \exp\Big(-\frac{\lambda}{2}\Big(\frac{1}{\langle \nu(s), p(s)\rangle^2} + |u|^2 + v^2|p(s)| + 2v \Big)\Big) \,\mathrm{d}u \,\mathrm{d}v \,\mathrm{d}s \,. \end{aligned}$$

We first perform the integration in u, obtaining the upper bound

$$\begin{aligned} &(1+\epsilon) \frac{|\langle p, p''\rangle|}{|p|^2|p'|}(1/2) \frac{(2\pi)^{(d-2)/2}}{\lambda^{(d-2)/2}} \times \\ &\quad \int_{|s-1/2|\leqslant \eta} \int_{0 \leqslant v \leqslant \eta} \exp\Big(-\frac{\lambda}{2}\Big(\frac{1}{\langle \nu(s), p(s)\rangle^2} + v^2|p(s)|^2 + 2v \Big)\Big) \,\mathrm{d}s \,\mathrm{d}v \end{aligned}$$

To perform the integration in v, we write

$$\begin{aligned} &\exp\Big(-\frac{\lambda}{2}\big(v^2|p(s)|^2 + 2v \big)\Big) \\ &\qquad = \Big(\lambda\big(v|p(s)|^2 + 2 \big)\Big)^{-1} \frac{\mathrm{d}}{\mathrm{d}v} \exp\Big(\frac{\lambda}{2}\big(v^2|p(s)|^2 + 2v \big)\Big) \end{aligned}$$

and integrate by parts. We obtain

$$\int_{0 \leqslant v \leqslant \eta} \exp\Big(-\frac{\lambda}{2}\big(v^2|p(s)|^2 + 2v \big)\Big) \,\mathrm{d}v \sim \frac{1}{\lambda}$$

as λ tends to infinity. This yields the upper bound

$$(1+\epsilon)^2 \frac{|\langle p, p''\rangle|}{|p|^2|p'|}(1/2)\frac{(2\pi)^{(d-2)/2}}{\lambda^{d/2}}\int_s \exp\Big(-\frac{\lambda}{2}\langle \nu(s), p(s)\rangle^{-2}\Big)\mathrm{d}s\,.$$

To estimate this last integral boils down to using the classical Laplace method. We introduce

$$\delta = \langle \nu, p\rangle^2 = |p|^2 - \langle p, \tau\rangle^2\,.$$

We then have

$$\int_{|s-1/2|\leqslant\eta} \exp\Big(-\frac{\lambda}{2}\langle \nu(s), p(s)\rangle^{-2}\Big)\,\mathrm{d}s$$

$$= \exp\Big(-\frac{\lambda}{2}\delta(1/2)\Big)\int_{|h|\leqslant\eta}\exp\Big(-\frac{\lambda}{2}\frac{\delta(1/2)-\delta(1/2+h)}{\delta(1/2)\delta(1/2+h)}\Big)\,\mathrm{d}h\,. \quad (12.4.2)$$

We then need to obtain a Taylor expansion for $\delta(\cdot)$ near $1/2$. To this end, for any integer m, we define

$$S_m = \sum_{1\leqslant k\leqslant d}(-1)^k k^m\,.$$

Since

$$|p(s)|^2 = \frac{2}{\pi^2}\sum_{1\leqslant k\leqslant d}\frac{\sin^2(k\pi s)}{k^2} = \frac{1}{\pi^2}\sum_{1\leqslant k\leqslant d}\frac{1-\cos(2k\pi s)}{k^2}\,,$$

we easily obtain the derivatives $(\mathrm{d}^m/\mathrm{d}s^m)\big(|p(s)|^2\big)$ at $s = 1/2$, for $m = 0, 1, \ldots, 4$. Using Taylor's formula, we infer that

$$|p(1/2+h)|^2 = |p(1/2)|^2 + 2h^2 S_0 - \frac{2}{3}\pi^2 h^4 S_2 + O(h^5)\,.$$

We also have

$$\langle p, p'\rangle(s) = \frac{2}{\pi}\sum_{1\leqslant k\leqslant d}\frac{\sin(k\pi s)\cos(k\pi s)}{k} = \frac{1}{\pi}\sum_{1\leqslant k\leqslant d}\frac{\sin(2k\pi s)}{k}\,.$$

Hence, a simple calculation of the derivatives at $s = 1/2$ and an application of Taylor's formula give

$$\langle p, p'\rangle(1/2+h) = 2hS_0 - \frac{4}{3}\pi^2 h^3 S_2 + O(h^5)\,.$$

Finally,

$$|p'(s)|^2 = 2 \sum_{1\leqslant k\leqslant d} \cos^2(k\pi s) = d + \sum_{1\leqslant k\leqslant d} \cos(2k\pi s)\,,$$

from which we deduce $|p'(1/2)|^2 = d + S_0$, and

$$|p'(1/2+h)|^2 = |p'(1/2)|^2 - 2\pi^2 h^2 S_2 + O(h^4)\,.$$

Equipped with these expressions, a little algebra gives

$$\begin{aligned}\delta(1/2+h) = \delta(1/2) &+ 2S_0\Big(1 - \frac{2S_0}{|p'(1/2)|^2}\Big)h^2 \\ &+ \pi^2 S_2\bigg(-\frac{2}{3} - \frac{S_0}{|p'(1/2)|^2}\Big(-\frac{16}{3} + \frac{8S_0}{|p'(1/2)|^2}\Big)\bigg)h^4 + O(h^5)\,.\end{aligned}$$

The interesting fact is now that S_0 equals -1 if d is odd, and equals 0 if d is even. Thus, if d is odd,

$$\delta(1/2+h) = \delta(1/2) - 2\Big(1 + \frac{2}{|p'(1/2)|^2}\Big)h^2 + O(h^4)\,,$$

while if d is even,

$$\delta(1/2+h) = \delta(1/2) - \frac{2\pi^2 S_2}{3}h^4 + O(h^5)\,.$$

Notice that when d is even, S_2 is positive since

$$\begin{aligned}\sum_{1\leqslant k\leqslant 2m} (-1)^k k^2 = \sum_{1\leqslant k\leqslant m} (2k)^2 - (2k-1)^2 &= \sum_{1\leqslant k\leqslant m} (4k-1) \\ &= m(2m+1)\,.\end{aligned}$$

Hence, $\delta(s)$ is maximal at $s = 1/2$ whatever the parity of d is.

To conclude the proof of Theorem 12.4.1, assume first that d is odd. Then, for η small enough, (12.4.2) is less than

$$\begin{aligned}\int_{|h|\leqslant\eta} \exp\bigg(-\lambda\Big(1 + \frac{2}{|p'(1/2)|^2}\Big)\frac{h^2}{\delta(1/2)^2(1+\epsilon)^2}\bigg)\,\mathrm{d}h \\ \sim \frac{\sqrt{2\pi}(1+\epsilon)\delta(1/2)}{\sqrt{1+2|p'(1/2)|^{-2}}}\frac{1}{\sqrt{\lambda}}\,.\end{aligned}$$

All the arguments we used to obtain this upper bound can be used to obtain a lower bound, essentially by changing ϵ to $-\epsilon$. Since ϵ is arbitrary, combining all the estimates yields

$$P\{\sup_{0\leqslant s\leqslant 1} B_d(s) \geqslant t\} \sim \sqrt{\frac{d+1}{2\pi}}\frac{1}{t}\exp\Big(-\frac{t^2}{2\sigma_d^2}\Big)$$

When d is even, for η small enough, the integral in (12.4.2) is less than

$$\int_{|h|\leqslant\eta} \exp\Big(-\frac{\lambda}{2}\frac{2\pi^2 S_2}{3}\frac{h^4}{\delta(1/2)^2(1+\epsilon)^2}\Big)\,\mathrm{d}h \\ \sim \frac{\Gamma(1/4)}{\lambda^{1/4}}\Big(\frac{3}{S_2}\Big)^{1/4}\sqrt{\frac{(1+\epsilon)\delta(1/2)}{\pi}}\,.$$

It then follows that

$$P\{ \sup_{0\leqslant s\leqslant 1} B_d(s) \geqslant t \} \sim \frac{\sqrt{d}}{4\pi^{3/2}}\frac{\Gamma(1/4)}{\sigma_d}\Big(\frac{6}{d(d+1)}\Big)^{1/4}\frac{1}{\sqrt{t}}\exp\Big(-\frac{t^2}{2\sigma_d^2}\Big)\,.$$

This concludes the proof of Theorem 12.4.1. ∎

To conclude this section let us make a remark on σ_d^2. Recall that $\sum_{k\geqslant 1}(2k+1)^{-2} = \pi^2/8$. Consequently, $\lim_{d\to\infty}\sigma_d^{-2} = 2$, which is good, given (12.4.1). We can obtain a more precise estimate of σ_d^2. Since

$$\int_m^\infty \frac{\mathrm{d}x}{(2x+1)^2} \leqslant \sum_{k\geqslant m}\frac{1}{(2k+1)^2} \leqslant \int_m^\infty \frac{\mathrm{d}x}{(2x-1)^2}\,,$$

and

$$\int_m^\infty \frac{\mathrm{d}x}{(2x+1)^2} \leqslant \int_m^\infty \frac{\mathrm{d}x}{4x^2} \leqslant \int_m^\infty \frac{\mathrm{d}x}{(2x-1)^2}\,,$$

we have

$$\Big|\sum_{k\geqslant m}\frac{1}{(2k+1)^2} - \frac{1}{4m}\Big| \leqslant \int_m^\infty \frac{1}{(2x-1)^2} - \frac{1}{(2x+1)^2}\,\mathrm{d}x \\ = \frac{1}{2(4m^2-1)}\,.$$

Consequently,

$$\Big|\sigma_d^2 - \frac{1}{4} - \frac{1}{2\pi d}\Big| \leqslant \frac{1}{\pi^2(4d^2-1)}\,.$$

Though the constant in the exponential term in Theorem 12.4.1, namely $1/(2\sigma_d^2)$, has the right limit as d tends to infinity, we cannot take the limit of the polynomial term and recover (12.4.1). This is caused by the slow convergence of B_d to the Brownian bridge.

Again, the whole message is to be rather cautious with these approximations. Strange things may happen and further examination is certainly needed if they are to be used in serious applications.

12.5. Polar processes on boundary of convex sets.

Let C be a bounded convex set in $\mathbb{R}^d$, with nonempty interior. The purpose of this section is to construct a simple process on its boundary. This construction is suggested by Theorem 7.1 and the results of chapter 9.

We first need to make some remarks on densities proportional to e^{-I} with I strictly convex and α-homogeneous. Recall that Λ_c denotes the level line $I^{-1}(\{c\})$. Every nonzero point x of $\mathbb{R}^d$ can be written in a unique way as $x = s\lambda$ for some positive s and λ in Λ_1. In this (λ, s)-coordinate system, the measure e^{-I} can be rewritten as

$$s^{d-1}e^{-s^\alpha}|\mathrm{Proj}_{T_\lambda\Lambda_1^\perp}\lambda|\,\mathrm{d}s\,\mathrm{d}\mathcal{M}_{\Lambda_1}(\lambda)\,. \tag{13.5.1}$$

Conversely, such a measure corresponds to a log-concave and log-α-homogeneous measure on $\mathbb{R}^d$.

Going back to the convex set C given at the beginning of this section, assume that it also contains the origin. Let Λ be the polar reciprocal of ∂C. For λ in Λ and nonnegative s, define $I(s\lambda) = s^\alpha$. Then, $\Lambda = \Lambda_1$ for this specific function I. Equip $\mathbb{R}^d$ with the log-concave density proportional to (13.5.1), and let X be a random vector having this density. We can consider the process $p \in \partial C \mapsto \langle X, p\rangle \in \mathbb{R}$. We call this process a polar process on ∂C.

The tail distribution of its supremum is given by the following result.

12.5.1. THEOREM. *For the polar process on ∂C defined above, and $1/c = \int_{\mathbb{R}^d} e^{-I(x)}\,\mathrm{d}x$,*

$$P\{\sup_{p\in\partial C}\langle X, p\rangle \geqslant t\} \sim \frac{c}{\alpha}e^{-t^\alpha}t^{d-\alpha}\int_\Lambda |\mathrm{Proj}_{(T_\lambda\Lambda)^\perp}\lambda|\,\mathrm{d}\mathcal{M}_\Lambda(\lambda)$$

as t tends to infinity.

Proof. We apply Theorem 7.1. In view of section 9.1, the dominating manifold is Λ, of dimension $k = d-1$. Moreover, by construction, $\partial A_1 = \Lambda_1$. It follows from Theorem 7.1 that the tail equivalent is of the form

$$ce^{-t^\alpha}t^{d-\alpha}\int_\Lambda \frac{\mathrm{d}\mathcal{M}_\Lambda}{|\mathrm{D}I|}\,.$$

To calculate $\mathrm{D}I$, notice that it is the outward normal to Λ. Its norm is obtained through the identity

$$\frac{\mathrm{d}}{\mathrm{d}s}I(\lambda s)\Big|_{s=1} = \mathrm{D}I(\lambda)\cdot\lambda = \frac{\mathrm{d}}{\mathrm{d}s}\big(s^\alpha I(\lambda)\big)\Big|_{s=1} = \alpha I(\lambda) = \alpha\,.$$

It implies

$$|\mathrm{D}I(\lambda)| = \alpha/|\mathrm{Proj}_{(T_\lambda\Lambda)^\perp}\lambda|\,,$$

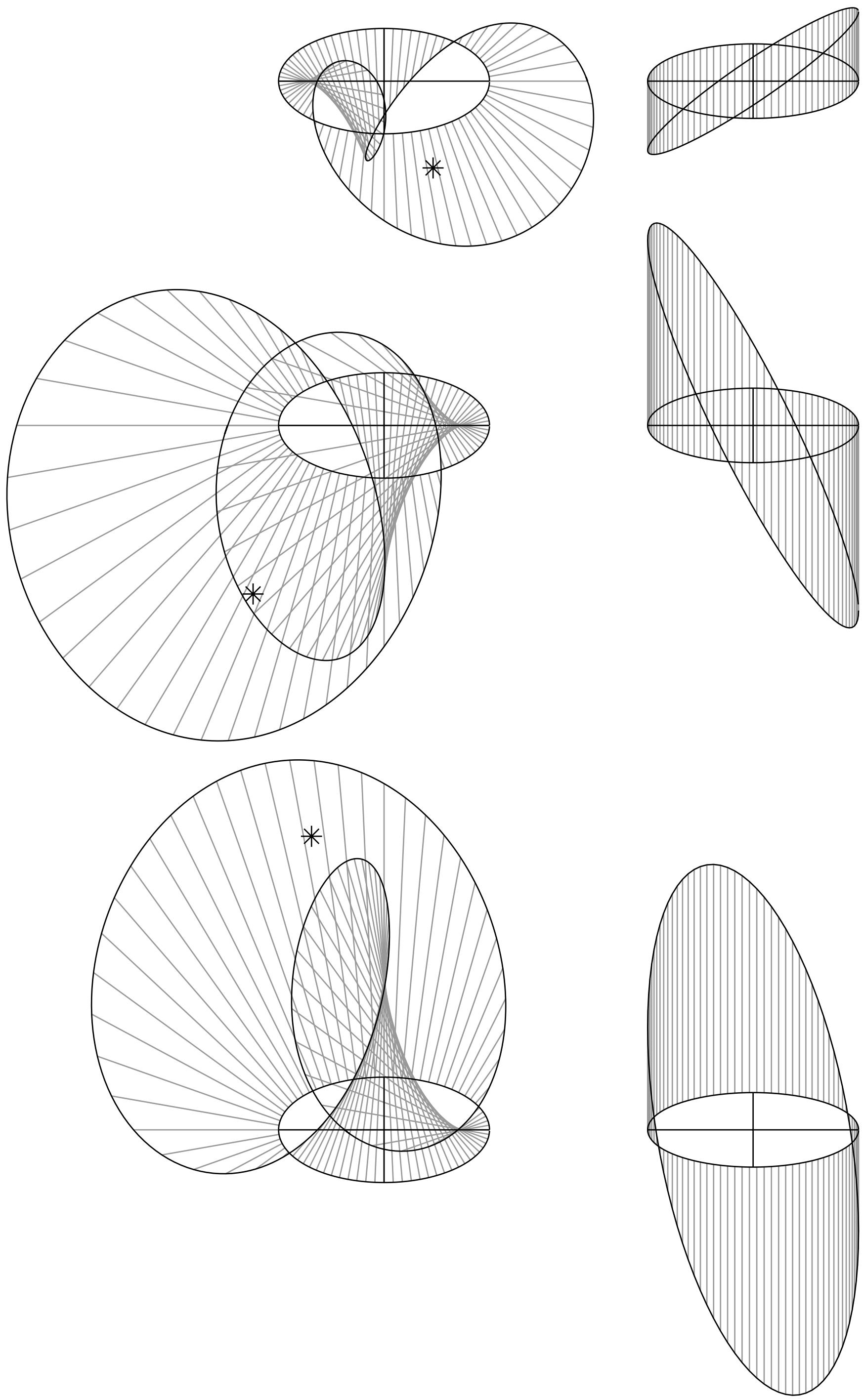
and the result follows. ■

The correspondence between points p on ∂C and points λ on the polar reciprocal allows us to write $|\mathrm{Proj}_{(T_\lambda\Lambda)^\perp}\lambda| = 1/|p|$. Thus the integral in Theorem 12.5.1 can also be rewritten as

$$\int_\Lambda \frac{1}{|p(\lambda)|}\mathrm{d}\mathcal{M}_\Lambda(\lambda)\,,$$

which is a nice formula.

The following pictures are three simulations of a polar process on the ellipsoid $x^2 + 2y^2 = 1$ in $\mathbb{R}^2$, and $\alpha = 1$. For the pictures on the left, at every point p of ∂C, we draw a segment in the normal direction to $T_p\partial C$, with length equal to the value of the process at p. The random point X is indicated by the star $\ast$. On the right hand side pictures, the process is represented as a curve held over the ellipsoid, the height of the curve being the value of the process; these are orthographic projections.



Notes

I do not think the results in this chapter are too serious! However, I believe the message of caution that some carry is important. In a different vein, Feynman supposedly said that when you have a theory, you should show where it works and where it does not. That may be the point of section 12.2.

Section 12.5 is just a way to generate pretty pictures of random caustics. I do not know any applications of this construction.

Appendix 1. Gaussian and Student tails

This appendix collects a few standard results related to the tails of the Gaussian and Student distributions. We write

$$\underline{s}_\alpha(x) = K_{s,\alpha}\Big(1+\frac{x^2}{\alpha}\Big)^{-(\alpha+1)/2}, \qquad x \in \mathbb{R}\,,\ \alpha > 0$$

for the Student density (with $\alpha+1$ degrees of freedom), the constant $K_{s,\alpha}$ ensuring that s_α integrates to 1 over the real line. We denote by

$$\underline{S}_\alpha(x) = \int_{-\infty}^{x} s_\alpha(y)\mathrm{d}y$$

the Student cumulative distribution function. Its tail is given in the following result.

A.1.1. LEMMA. *We have*

$$1-\underline{S}_\alpha(x) = \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{x^\alpha} + O\Big(\frac{1}{x^{\alpha+2}}\Big) \qquad \textit{as } x \to \infty\,.$$

Proof. Notice that as y tends to infinity,

$$\begin{aligned}\frac{1}{\Big(1+\frac{y^2}{\alpha}\Big)^{(\alpha+1)/2}} - \frac{1}{\Big(\frac{y^2}{\alpha}\Big)^{(\alpha+1)/2}} &= \frac{1-\Big(\frac{\alpha}{y^2}+1\Big)^{(\alpha+1)/2}}{\Big(1+\frac{y^2}{\alpha}\Big)^{(\alpha+1)/2}} \\ &= O\Big(\frac{1}{y^{\alpha+3}}\Big)\,.\end{aligned}$$

Consequently, as x tends to infinity,

$$\begin{aligned}\frac{1-S_\alpha(x)}{K_{s,\alpha}} = \int_x^\infty \frac{\mathrm{d}y}{\Big(1+\frac{y^2}{\alpha}\Big)^{\frac{\alpha+1}{2}}} &= \int_x^\infty \frac{\alpha^{(\alpha+1)/2}}{y^{\alpha+1}} + O\Big(\frac{1}{y^{\alpha+3}}\Big)\mathrm{d}y \\ &= \frac{\alpha^{(\alpha-1)/2}}{x^\alpha} + O\Big(\frac{1}{x^{\alpha+2}}\Big)\end{aligned}$$

∎

Let us now consider a Student-like cumulative distribution function S_α, i.e., such that

$$S_\alpha(-x) \sim 1 - S_\alpha(x) \sim \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{x^\alpha},$$

as x tends to infinity, and where the constant $K_{s,\alpha}$ can be any fixed positive number. We can obtain an asymptotic formula for the high quantiles.

A.1.2. LEMMA. *The following holds,*

$$(1 - S_\alpha)^{\leftarrow}(u) \sim \frac{K_{s,\alpha}^{1/\alpha}\,\alpha^{(\alpha-1)/2\alpha}}{u^{1/\alpha}} \qquad \text{as } u \to 0\,.$$

Proof. Let x tends to infinity and u tends to 0 such that $u = 1 - S_\alpha(x)$, that is $x = (1 - S_\alpha)^{\leftarrow}(u)$. From the Student-like tail, we infer that

$$u \sim \frac{K_{s,\alpha}\alpha^{(\alpha-1)/2}}{x^\alpha},$$

which is the result. ■

We can obtain similar results for the Gaussian distribution with cumulative distribution function

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\mathrm{d}y\,.$$

A.1.3. LEMMA. *We have*

$$1 - \Phi(x) = \frac{e^{-x^2/2}}{x\sqrt{2\pi}}\big(1 + o(1)\big) \qquad \text{as } x \to \infty\,.$$

Proof. Integrate by parts to obtain

$$\begin{aligned}\sqrt{2\pi}\big(1 - \Phi(x)\big) &= \int_x^\infty \frac{1}{y} y e^{-y^2/2}\mathrm{d}y \\ &= \frac{e^{-x^2/2}}{x} - \int_x^\infty \frac{e^{-y^2/2}}{y^2}\mathrm{d}y \\ &= \frac{e^{-x^2/2}}{x} - \frac{e^{-x^2/2}}{3x^3} + \int_x^\infty \frac{e^{-y^2/2}}{4y^4}\mathrm{d}y\,.\end{aligned}$$

Consequently,

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}}\Big(\frac{1}{x}-\frac{1}{3x^2}\Big) \leqslant 1-\Phi(x) \leqslant \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\frac{1}{x}\,,$$

and the result follows. ■

We can now obtain an asymptotic approximation for high quantiles. The second statement in the following lemma is instrumental in the sequel.

A.1.4. LEMMA. *We have*

$$\Phi^{\leftarrow}(1-u)=\sqrt{2\log 1/u}-\frac{\log\log 1/u}{2\sqrt{2\log 1/u}}-\frac{\log(2\sqrt{\pi})}{\sqrt{2\log 1/u}}\big(1+o(1)\big)$$

as u tends to 0*, and*

$$\Phi^{\leftarrow}(1-u)^2=2\log 1/u-\log\log 1/u-2\log(2\sqrt{\pi})+o(1) \qquad \textit{as } u\to 0\,.$$

Proof. As in the proof of Lemma A.1.2, we start with the equality $u=1-\Phi(x)$, that is $x=(1-\Phi)^{\leftarrow}(u)$. We consider x tending to infinity, or equivalently u converging to 0. Lemma A.1.3 implies

$$\log u=-\frac{x^2}{2}-\log x-\log\sqrt{2\pi}+o(1) \qquad \text{as } x\to\infty\,,\ u\to 0\,. \quad \text{(A.1.1)}$$

Consequently, $x=\sqrt{2\log 1/u}(1+x_1)$ with x_1 tending to 0 as u tends to 0. Then (A.1.1) yields

$$\begin{aligned}\log u &= \Big(-\log\frac{1}{u}\Big)(1+x_1)^2-\frac{1}{2}\log\log\frac{1}{u}-\log(2\sqrt{\pi})\\ &\qquad\qquad -\log(1+x_1)+o(1)\\ &= \log u-2x_1\log\frac{1}{u}-x_1^2\log\frac{1}{u}-\frac{1}{2}\log\log\frac{1}{u}\\ &\qquad\qquad -\log(2\sqrt{\pi})+o(1)\,. \quad \text{(A.1.2)}\end{aligned}$$

We can then calculate

$$x_1=-\frac{\log\log 1/u}{4\log 1/u}(1+x_2)$$

with x_2 tending to 0 with u. But now (A.1.2) implies

$$\begin{aligned}0 &= \frac{1+x_2}{2}\log\log 1/u - \frac{1}{16}\frac{(\log\log 1/u)^2}{\log 1/u}(1+x_2)^2 \\ &\qquad\qquad - \frac{1}{2}\log\log 1/u - \log(2\sqrt{\pi}) + o(1) \\ &= \frac{x_2}{2}\log\log\frac{1}{u} - \log(2\sqrt{\pi}) + o(1)\,,\end{aligned}$$

and thus, as u tends to 0,

$$x_2 = \frac{2\log(2\sqrt{\pi})}{\log\log 1/u}\bigl(1+o(1)\bigr)\,.$$

Consequently, gathering every piece yields

$$\begin{aligned}x &= \sqrt{2\log 1/u} - \frac{1}{2\sqrt{2}}\frac{\log\log 1/u}{\sqrt{\log 1/u}}(1+x_2) \\ &= \sqrt{2\log 1/u} - \frac{1}{2\sqrt{2}}\frac{\log\log 1/u}{\sqrt{\log 1/u}} - \frac{1}{\sqrt{2}}\frac{\log(2\sqrt{\pi})}{\sqrt{\log 1/u}}\bigl(1+o(1)\bigr)\end{aligned}$$

as u tends to 0, which is the desired expression for $\Phi^{\leftarrow}(1-u)$. Square it to obtain that for $\Phi^{\leftarrow}(1-u)^2$. ∎

We can now obtain an asymptotic expansion for $\Phi^{\leftarrow}\circ S_\alpha$.

A.1.5. LEMMA. *We have*

$$\begin{aligned}\Phi^{\leftarrow}\circ S_\alpha(x) = \sqrt{2\alpha\log x} - \frac{\log\log x}{2\sqrt{2\alpha\log x}} - \frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{\sqrt{2\alpha\log x}} \\ + o\Bigl(\frac{1}{\sqrt{\log x}}\Bigr)\end{aligned}$$

as x tends to infinity. Consequently,

$$\Phi^{\leftarrow}\circ S_\alpha(x)^2 = 2\alpha\log x - \log\log x - 2\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi}) + o(1)$$

as x tends to infinity.

Proof. From Lemma A.1.4, we deduce that

$$\begin{aligned}\Phi^{\leftarrow}\circ S_\alpha(x)^2 = -2\log\bigl(1-S_\alpha(x)\bigr) - \log\bigl(-\log(1-S_\alpha(x)\bigr) \\ - 2\log(2\sqrt{\pi}) + o(1)\end{aligned}$$

as x tends to infinity. But Lemma A.1.1 implies

$$-\log\bigl(1-S_\alpha(x)\bigr)=\alpha\log x-\log(K_{s,\alpha}\alpha^{(\alpha-1)/2})+o(1)$$

as x tends to infinity. Consequently,

$$\begin{aligned}\Phi^{\leftarrow}\circ S_\alpha(x)^2=2\alpha\log x-2\log(K_{s,\alpha}\alpha^{(\alpha-1)/2})-\log\log x-\log\alpha\\ -2\log(2\sqrt{\pi})+o(1)\end{aligned}$$

as x tends to infinity, which is the second expansion in Lemma A.1.5. Taking the square root yields the first assertion since

$$\begin{aligned}&\Phi^{\leftarrow}\circ S_\alpha(x)\\ &\quad=\sqrt{2\alpha\log x}\Bigl(1-\frac{\log\log x}{2\alpha\log x}-\frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{\alpha\log x}+o\Bigl(\frac{1}{\log x}\Bigr)\Bigr)^{1/2}\\ &\quad=\sqrt{2\alpha\log x}\Bigl(1-\frac{\log\log x}{4\alpha\log x}-\frac{\log(K_{s,\alpha}\alpha^{\alpha/2}2\sqrt{\pi})}{2\alpha\log x}+o\Bigl(\frac{1}{\log x}\Bigr)\Bigr)\,.\end{aligned}$$ ∎

We can also obtain an estimate of $S_\alpha^{\leftarrow}\circ\Phi(x)$.

A.1.6. LEMMA. *We have*

$$\log S_\alpha^{\leftarrow}\circ\Phi(x)=\frac{x^2}{2\alpha}+\frac{1}{\alpha}\log x+\frac{1}{\alpha}\log(K_{s,\alpha}\alpha^{\frac{\alpha-1}{2}}\sqrt{2\pi})+o(1)$$

as x tends to infinity.

Proof. Combine Lemma A.1.2 and A.1.3 to obtain

$$\begin{aligned}S_\alpha^{\leftarrow}\circ\Phi(x)&=(1-S_\alpha)^{\leftarrow}\circ(1-\Phi)(x)\\ &\sim K_{s,\alpha}^{1/\alpha}\alpha^{(\alpha-1)/2\alpha}x^{1/\alpha}(2\pi)^{1/2\alpha}e^{x^2/2\alpha}\end{aligned}$$

as x tends to infinity. The result follows by taking the logarithm. ∎

In the case of an exact Student distribution, the term $o(1)$ in Lemma A.1.6 is actually $O(x^{-1/\alpha}e^{-x^2/\alpha})$. In this special case, the approximation has a terrific accuracy as x tends to infinity!

Appendix 2.
Exponential map

The purpose of this appendix is to state and prove the following proposition. It gives a bound on the error committed by linearizing the exponential map over the level set of a function.

A.2.1. PROPOSITION. *Let $I:\mathbb{R}^d\to\mathbb{R}$ be a smooth smooth function, and let c be a regular value of I. Then $\Lambda_c = I^{-1}(c)$ is a smooth manifold of $\mathbb{R}^d$. Let*

$$M=\sup\Big\{\frac{|\mathrm{D}^2 I(p)|}{|\mathrm{D}I(p)|}\,:\,p\in\mathbb{R}^d\,,\;I(p)=c\Big\}\,.$$

Then, for any p in Λ_c, any u in $T_p\Lambda_c$ with $|u|<1/4M$,

$$|\exp_p(u)-p-u|\leqslant M|u|^2\,.$$

Proof. Let p be a point in Λ_c and v a unit tangent vector to Λ_c at p. We denote by $\gamma(s)=\exp_p(sv)$ the geodesic starting at p in the direction u on Λ_c. This parameterization is by arc length. Recall that $N=\mathrm{D}I/|\mathrm{D}I|$ is an outward unit normal vector field to the level set of I. Since

$$\mathrm{d}N(p)=\Big(\mathrm{Id}-\frac{1}{2}NN^{\mathrm{T}}\Big)\frac{\mathrm{D}^2 I}{|\mathrm{D}I|}(p)\,,$$

we have $\|\mathrm{d}N(p)\|\leqslant M$. Consequently,

$$\big|N\big(\gamma(s)\big)-N(p)\big|=\Big|\int_0^s \mathrm{d}N\big(\gamma(r)\big)\cdot\gamma'(r)\mathrm{d}r\Big|\leqslant sM\,.$$

The geodesic $\gamma(\cdot)$ is characterized by the parallel transport of its tangent vectors along $\gamma(s)$ which can be written as

$$\big[\mathrm{Id}-NN^{\mathrm{T}}\big(\gamma(s)\big)\big]\gamma''(s)=\mathrm{Proj}_{T_{\gamma(s)}\Lambda_c}\gamma''(s)=0\,.$$

Any w in $T_p\Lambda_c$ is orthogonal to $M(p)$. Consequently,

$$\langle w,\gamma''(s)\rangle=\big\langle w,\big[NN^{\mathrm{T}}\big(\gamma(s)\big)-NN^{\mathrm{T}}(p)\big]\cdot\gamma''(s)\big\rangle\,.$$

By duality,

$$\begin{aligned}
|\mathrm{Proj}_{T_p\Lambda_c}\gamma''(s)| &\leqslant \big\|NN^{\mathrm{T}}\big(\gamma(s)\big) - NN^{\mathrm{T}}(p)\big\|\,|\gamma''(s)| \\
&\leqslant 2\big|N\big(\gamma(s)\big) - NN^{\mathrm{T}}(p)\big|\,|\gamma''(s)| \\
&\leqslant 2sM|\gamma''(s)|\,.
\end{aligned}$$

Moreover, since $\big\langle N\big(\gamma(s)\big)\,,\gamma'(s)\big\rangle = 0$ and γ is parametrized by arc length,

$$\big|\big\langle N\big(\gamma(s)\big)\,,\gamma''(s)\big\rangle\big| = \big|-\big\langle \mathrm{d}N\big(\gamma(s)\big)\cdot\gamma'(s)\,,\gamma'(s)\big\rangle\big| \leqslant M\,.$$

It follows that

$$\begin{aligned}
|\mathrm{Proj}_{(T_p\Lambda_c)^\perp}\gamma''(s)| &= |\langle N(p)\,,\gamma''(s)\rangle| \\
&\leqslant \big|N(p) - N\big(\gamma(s)\big)\big|\,|\gamma''(s)| + M \\
&\leqslant sM|\gamma''(s)| + M\,.
\end{aligned}$$

Consequently, we have the inequality

$$\begin{aligned}
|\gamma''(s)|^2 &= |\mathrm{Proj}_{T_p\Lambda_c}\gamma''(s)|^2 + |\mathrm{Proj}_{(T_p\Lambda_c)^\perp}\gamma''(s)|^2 \\
&\leqslant 4s^2M^2|\gamma''(s)|^2 + M^2(s|\gamma''(s)| + 1)^2\,.
\end{aligned}$$

On the range $s \leqslant 1/(4M)$, this inequality implies

$$|\gamma''(s)| \leqslant \frac{|\gamma''(s)|^2}{4} + \frac{|\gamma''(s)|^2}{8} + 2M^2\,,$$

that is

$$|\gamma''(s)|^2 \leqslant 16M^2/5 \leqslant 4M^2\,.$$

Consequently, since $\gamma'(0) = v$,

$$|\gamma'(s) - v| = \Big|\int_0^s \gamma''(t)\mathrm{d}t\Big| \leqslant 2Ms\,.$$

The result follows, since

$$|\gamma(s)-p-sv| = \Big|\int_0^s \big(\gamma'(t)-v\big)\mathrm{d}t\Big| \leqslant \int_0^s 2Mt\mathrm{d}t = Ms^2 = M|sv|^2\,. \quad \blacksquare$$

References

Adler, R.J. (2000). On excursion sets, tube formulas and maxima of random fields, *Ann. Appl. Probab.*, 10, 1–74.

Andrews, G.E., Askey, R., Roy, R. (1999). *Special Functions*, Cambridge University Press.

Balkema, A.A., Klüppelberg, C., Resnick, S.I. (1999). Limit laws for exponential families, *Bernoulli*, 5, 951–968.

Barndorff-Nielsen, O.E. (1978). *Information and Exponential Families in Statistical Theory*, Wiley.

Barndorff-Nielsen, O.E., Klüppelberg, C. (1999). Tail exactness of multivariate saddlepoint approximations, *Scand. J. Statist.* 26 (1999), 253–264.

Barndorff-Nielsen, O.E. (1990). Approximation of interval probabilities, *J. Roy. Statist. Soc.*, B, 52, 485–496.

Barbe, Ph., Broniatowski, M. (200?). A sharp Petrov type large deviation formula, submitted.

Billinsley, P. (1968). *Convergence of Probability Measures*, Wiley.

Bingham, N.A., Goldie, C.M., Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

Bishop, R.L. (1977). Decomposition of cut loci, *Proc. Amer. Math. Soc.*, 65, 133–136.

Bleistein, N., Handelman, R. (1975). *Asymptotic Expansions of Integrals*, Holt, Reinhartm Winston, New York. Republished by Dover in 1986.

Bobkov, S., Ledoux, M. (1997). Poincaré's inequality and Talagrand's concentration phenomenon for the exponential distribution, *Probab. Theory Rel. Fields*, 107, 383–400.

Bobokov, S.G., Ledoux, M. (2000). From Brunn-Minkowski to Bracamp-Lieb and to logarithmic Sobolev inequalities, *Geom. Funct. Anal.*, 10, 1028–1052.

Bolthausen, E. (1993). Stochastic processes with long range interactions of the paths, *Doeblin and Modern Probability*, Bauberen,

1991, pp.297–310, *Contemp. Math.*, 149, Amer. Math. Soc., Providence, R.I.

Breitung, K. (1994). *Asymptotic Approximations for Probability Integrals*, *Lecture Notes in Mathematics*, 1592, Springer.

Breitung, K., Hohenbichler, M. (1989). Asymptotic approximations for multivariate integrals with an application to multinormal probabilities, *J. Multivariate Anal.*, 30, 80–97.

Breitung, K., Richter, W.-D. (1996). A geometric approach to an asymptotic expansion for large deviation probabilities of Gaussian random vectors, *J. Multivariate Anal.*, 58, 1–20.

Broniatowski, M., Fuchs, A. (1995). Tauberian theorems, Chernoff inequality, and the tail behavior of finite convolutions of distribution functions, *Adv. Math.*, 116, 12–33.

Brockwell, P.J., Davis, R.A. (1987). *Time Series: Theory and Methods*, Springer.

Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics.

Bruce, J.W., Giblin, P.J. (1992). *Curves and Singularities*, 2nd ed., Cambridge University Press.

Chavel, I. (1996). *Riemannian Geometry: A Modern Introduction*, Cambridge University Press.

Combet, E. (1982). *Intégrales Exponentielles*, *Lecture Notes in Mathematics*, 937, Springer.

Cox, D., Little, J., O'Shea, D. (1992). *Ideals, Varieties, and Algorithms*, 2nd ed., Springer.

Cox, D., Little, J., O'Shea, D. (1998). *Using Algebraic Geometry*, Springer.

Csiszár, I. (1984). Sanov property, generalized I-projection and a conditional limit theorem, *Ann. Probab.*, 12, 768–793.

Davis, R., Resnick, S.I. (1986). Limit theory for the sample covariance and correlation functions of moving averages, *Ann. Statist.*, 14, 533-558.

De Bruijn, N.G. (1958). *Asymptotic Methods in Analysis*, North Holland, Republished by Dover in 1981.

Diebolt, J., Posse, Ch. (1996). On the density of the maximum of smooth Gaussian processes. *Ann. Probab.*, 24, 1104–1129.

Do Carmo, M. P. (1976). *Differential Geometry of Curves and Surfaces*, Prentice Hall.

Do Carmo, M.P. (1992). *Riemannian Geometry*, Birkhauser.

Dembo, A., Zeitouni, O. (1993). *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers, Boston, MA.

Dupuis, P., Ellis, R.S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley.

Federer, H. (1969). *Geometric Measure Theory*, Springer.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd ed., Wiley.

Fernique, X. (1970). Intégrabilité des vecteurs gaussiens, *C.R. Acad. Sci. Paris*, 270, 1698-1699.

Günther, P. (1960). Einige Sätze über das Volumenelement eines Riemannschen Raumes, *Pub. Math. Deberecen*, 7, 78–93.

Hasanis, T., Koutroufiotis, D. (1985). The characteristic mapping of a reflector. *J. Geom.*, 24, 131–167.

Hlawka, E., Schoißengeir, J., Taschner, R. (1991). *Geometric and Analytic Number Theory*, Springer.

Howard, R. (1993). *The Kinematic Formula in Riemannian Homogeneous Spaces*, *Memoires of the Amer. Math. Soc.*, 509.

Hwang, C.-R. (1980). Laplace's method revisited: weak convergence of probability measures, *Ann. Probab.*, 8, 1177–1182.

Imhoff, J.P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, 419–426.

Jackson, D. (1911). Ueler eine trigonometrishe Summe, *Rend. Circ. Mat. Palermo*, 32, 257–262.

Jensen, J.L. (1995). *Saddlepoint Approximations*, Oxford University Press.

Johnson, N.L., Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley.

Klingenberg, W. (1959). Contribution to Riemannian geometry in the large, *Ann. of Math.*, 69, 654–666.

Knapp, A.W. (1996). *Lie Groups Beyond an Introduction*, Birkhauser.

Landau, H.J., Shepp, L.A. (1970). On the supremum of Gaussian processes, *Sankyà*, A32, 369–378.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer.

LeCam, L., Yang, G.L. (1990). *Asymptotics in Statistics, Some Basic Concepts*, Springer.

Ledoux, M. (1996). *Isoperimetry and Gaussian Analysis*, in *Lectures on Probability Theory and Statistics, Ecole d'été de Probabilités de Saint Flour*, XXIV, 1994, P. Bernard Ed., *Lecture Notes in Mathematics*, 1648, Springer.

Lieb, E., Loss, M. (1997). *Analysis*, American Mathematical Society.

Marinov, M.S. (1980). Invariant volumes of compact groups, *J. Phys.*, A, 13, 3357–3366.

Mehta, M.L. (1991). *Random Matrices*, 2nd edition, Academic Press.

McCleary, J. (1974). *Geometry from a Differentiable Viewpoint*, Cambridge University Press.

Mijnheer, J. (1997a). Asymptotic inference for AR(1) processes with (nonnormal) stable errors, *J. Math. Sci.*, 83, 401–406.

Mijnheer, J. (1997b). Asymptotic inference for AR(1) processes with (nonnormal) stable errors, III, *Comm. Statist. Stochastic Models*, 13, 661–672.

Mijnheer, J. (1997c). Asymptotic inference for AR(1) processes with (nonnormal) stable errors, IV, A note on the case of a negative unit root, *J. Math. Sci.*, 92, 4035–4037.

Morgan, F. (1988). *Geometric Measure Theory, a Beginner's Guide*, Academic Press.

Morgan, F. (1992). *Riemannian Geometry, a Beginner's Guide*, Peters Ltd.

Olver, F.W.J. (1974). *Asymptotics and Special Functions*, Academic Press.

Piterbarg, V.V. (1996). *Asymptotic methods in the theory of Gaussian processes and fields, Translated from the Russian by V. V. Piterbarg*, Amer. Math. Soc.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer.

Press, W.H., Teukolsky, S.A., Vettering, N.H., Flannery, B.P. (1996). *Numerical Recipes*, Cambridge University Press.

Rausch, H.E. (1951). A contribution to differential geometry in the large, *Ann. of Math.*, 54, 38–55.

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer.

Rosiński, J., Woyczyński, W.A. (1987). Multilinear forms in Pareto-like random variables and product random measures, *Colloquium Mathematicum*, 305–313.

Schneider, R. (1993). *Convex Bodies: the Brunn-Minkowski Theory*, Cambridge University Press.

Skorokhod, A.V. (1956). Limit theorem for stochastic processes, *Theor. Probab. Appl.*, 1, 261–290.

Spivak,M. (1970). *A Comprehensive Introduction to Differential Geometry*, Publish or Perish.

Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces, *Publ. Math. IHES*, 81, 73–205.

Weyl, H. (1939). On the volume of tubes, Proc. Amer. Math. Soc., 65, 461–472.

Notation

This index has 3 parts. The first one concerns general notation, that is used throughout the book, eventually with variations in the arguments in each example.

The second part contains the notation introduced in chapters 2–5. This is used almost throughout the book, but in examples treated in chapters 6–12 is given a specialized meaning.

The third part contains the notation with the concrete meaning given through chapters 6–12.

Some notation that are used very locally does not appear in the index.

General notation.

$\sharp$	the cardinal of a set, as in $\sharp\{1,2,3\} = 3$.
d	integration element, as in $\mathrm{d}t$, $\mathrm{d}\mathcal{M}_M$, $\mathrm{d}x$.
d	dimension of the underlying space, $\mathbb{R}^d$.
D, D^2, ...	Gradient, Hessian, and higher order differentials.
$\exp_p(\cdot)$	exponential map at p (on a Riemannian manifold).
$\mathcal{M}_M$	Riemannian measure of a manifold M.
$.^\perp$	orthocomplement.
$\mathrm{inj}_M(p)$	radius of injectivity of p in the manifold M.
$K_M(x,y)$	sectional curvature of the manifold M along the tangent vector fields x and y.
$\lambda_{\min}(M)$, $\lambda_{\max}(M)$	smallest and largest eigenvalue of a matrix M.
∂A	boundary of a set.
$\partial f(\cdot)/\partial u$	partial differentiation of a function.
$\Pi_{M,p}$	second fundamental form of the manifold M at p.
$\mathbb{R}$, $\mathbb{R}^d$	set of real numbers, the Euclidean d dimensional space.
Ricc	Ricci tensor of the level lines of I.
S_n	unit sphere centered at the origin, of dimension n, that is the boundary of the unit ball in $\mathbb{R}^{n+1}$.
$S_V(x,r)$	ball centered at x, of radius r, in the vector space V.
T_pM	tangent space of the manifold M at p.
N_pM	normal space of the manifold M at p.
$\lvert S\rvert$	Lebesgue measure of the set S.
$\omega_n = \pi^n/\Gamma((n/2)+1)$	volume of the unit ball of $\mathbb{R}^n$.

Notation from chapters 2–5.

Notation from chapters 6–12.

Chapter 6

Chapter 7

Chapter 8

Chapter 9

Chapter 10

Chapter 11

Chapter 12

Postface

This book is the first of a larger project that I may try to complete. A second volume should be devoted to the asymptotic analysis of multivariate integrals over small wedges and their applications. A third one should extend some of the results of the first two volumes to the infinite dimensional setting, where there are some potentially amazing applications in the study of stochastic processes.

June 2001 Philippe Barbe